

Aus der Radiologischen Universitätsklinik (Department) Tübingen  
Abteilung Diagnostische und Interventionelle Radiologie  
Ärztlicher Direktor: Professor Dr. C. D. Claussen

**Entwicklung einer allgemeingültigen Standardprozedur  
zur Überprüfung der Gebrauchstauglichkeit  
medizintechnischer Produkte**

Inaugural-Dissertation zur Erlangung des Doktorgrades der  
Humanwissenschaften

der Medizinischen Fakultät der  
Eberhard Karls Universität zu Tübingen

vorgelegt von

**Dirk Michael Büchel**

**-2010-**

Dekan: Professor Dr. I. Autenrieth

1. Berichterstatter: Professor Dr. C. D. Claussen

2. Berichterstatter: Professor. Dr. G. Zimmermann

Diese Arbeit widme ich meinem Vater

# Inhaltsverzeichnis

<b><u>INHALTSVERZEICHNIS</u></b>	<b><u>2</u></b>
<b><u>1 EINLEITUNG</u></b>	<b><u>4</u></b>
1.1 Vorgehensweise und Aufbau der Arbeit	5
1.2 Theoretische Grundlagen	8
1.2.1 Mensch-Maschine-Kommunikation	8
1.2.2 Mensch-Maschine-Kommunikation in der Medizin	18
<b><u>2 ERGONOMIE UND GEBRAUCHSTAUGLICHKEIT</u></b>	<b><u>31</u></b>
2.1 Grundlegende Begriffe und Konzepte	31
2.1.1 Begriffsdefinition und Abgrenzung der Ergonomie und der Gebrauchstauglichkeit	31
2.1.2 Anthropometrie	35
2.1.3 Gebrauchstauglichkeit – Aber wie?	36
2.1.4 Entwicklungsprozess der Gebrauchstauglichkeit	37
2.2 Ergonomie und Gebrauchstauglichkeit in der Medizin	40
2.2.1 Notwendigkeit für Gebrauchstauglichkeit in der Medizin	42
2.2.2 Gründe für eine Gebrauchstauglichkeitsbetrachtung	47
2.2.3 Der Gebrauchstauglichkeitsentwicklungsprozess in der Medizin	56
2.3 Gesetze, Normen und Richtlinien	58
2.3.1 Gesetze	58
2.3.2 Normen	66
2.3.3 Diskussion der Relevanz der Gesetzgebung und Normen für diese Arbeit	71
2.4 Prüfprozess der Gebrauchstauglichkeit	72
2.4.1 Methoden der Gebrauchstauglichkeitsevaluation	76
2.4.2 Methodenvergleich	96
<b><u>3. ENTWICKLUNG EINER METHODE ZUR ÜBERPRÜFUNG DER GEBRAUCHSTAUGLICHKEIT MEDIZINISCHER PRODUKTE</u></b>	<b><u>108</u></b>
3.1 Ziele und Vorgehensweise	108
3.2 Ist-Situation: Diskussion der bestehenden Prüfverfahren im Hinblick auf die Standardprozedur	109
3.2.1 Ergonomieprüfung in der Medizin	109
3.2.2 Diskussion der Methoden und Rückschlüsse für eine Standardprozedur	121
3.3 Anforderungen an die Methode	123
<b><u>4. ENTWICKLUNG DER EVALUATIONSMETHODE</u></b>	<b><u>127</u></b>
4.1 Material und Methode	127
4.1.1 Anforderungsanalyse für medizinische Produkte	128
4.1.2 Prüfkriterien – Messung der Gebrauchstauglichkeit	137
4.1.3 Die Evaluatoren	143
4.1.4 Bestimmung der Stichprobe für den Nutzertest	146
4.1.5 Szenarien und Testaufgaben	149
4.1.6 Testaufbau und Testumgebung	153
4.1.7 Report	155
4.1.8 Qualität des standardisierten Verfahrens	157
4.2 Iterative Entwicklung und Evaluation der „Standardprozedur zur Prüfung der Gebrauchstauglichkeit medizinischer Geräte“	162
4.2.1 Entwicklungsstudie I – Studie zur Ermittlung der Gebrauchstauglichkeit von drei Hochfrequenz-Chirurgiegeräten	162
4.2.2 Entwicklungsstudie II – Studie zur Ermittlung der Gebrauchstauglichkeit von MIC Handgriffen	173
4.2.3 Entwicklungsstudie III – Evaluation einer Spritzenpumpe	181
4.3 Ergebnisse	189

<b><u>5 GRUNDLAGEN UND RAHMENBEDINGUNGEN FÜR DIE ENTWICKLUNG EINER BEWERTUNGSMETHODE DER GEBRAUCHSTAUGLICHKEIT MEDIZINISCHER PRODUKTE</u></b>	<b>200</b>
5.1 Fehler und Fehlerklassifikation	202
5.1.1 Menschliche Fehler	204
5.1.2 Klassifizierungen menschlicher Fehler	205
5.1.3 Fehler der Mensch-Maschine-Kommunikation und Klassifizierung von Gebrauchstauglichkeitsproblemen	209
5.1.4 Klassifizierung von Fehlern in sicherheitskritischen Systemen	214
5.1.5 Klassifikation von Fehlern in der Medizin	216
5.2 Erstellung eines Fehlerklassifizierungssystems unter Berücksichtigung der Mensch-Maschine-Schnittstelle	223
5.3 Sicherheit und sicherheitsbezogenen Begriffe	224
5.3.1 Sicherheitskritische Systeme	228
5.4 Methoden zur Analyse der Sicherheit	229
5.4.1 Risikoanalyse	231
5.4.2 Risikobewertung	239
5.4.3 Vertretbarkeit des Risikos	245
5.5 Die Komponente Mensch im Risikomanagement	247
5.6 Sicherheit und Gebrauchstauglichkeit	251
5.7 Menschliche Fehler und Sicherheit	253
5.8 Fehlerbewertung in der Mensch-Maschine-Interaktion	254
5.8.1 Fehleridentifikation	255
5.8.2 Bewertung der Bedeutung eines Fehlers	256
5.8.3 Beurteilung nach Auftretenswahrscheinlichkeit	256
5.8.4 Beurteilung nach dem Schweregrad	257
5.8.5 Klassifizierung nach potentielltem Schaden	259
<b><u>6 ENTWICKLUNG DES BEWERTUNGSMAßSTABES ZUR BESTIMMUNG DER GEBRAUCHSTAUGLICH MEDIZINISCHER GERÄTE</u></b>	<b>263</b>
6.1 Verschiedene Bewertungskonzepte	263
6.1.1 Konzept 1 – 3D-Matrix	263
6.1.2 Konzept 2 - Klassifizierung	269
6.1.3 Konzept 3 - Nutzwertanalyse	270
6.1.4 Konzept 4 – Risikograph	272
6.2 Auswahl, Entwicklung und Evaluation der Bewertungs-methode	273
6.2.1 Entwicklungsstudie 1 – Auswahl der Methode	274
6.2.2 Entwicklungsstudie 2 - Evaluation der Bewertungsmethode	275
6.2.3 Entwicklungsstudie 3 - Bewertung der HF-Geräte	283
6.2.4 Entwicklungsstudie 4 - Bewertung der Spritzenpumpe	291
6.3 Ergebnis - Bewertungsmethode „UseProb“	295
6.4 Validierung - Finale Evaluation	297
<b><u>7 ABSCHLIEBENDE DISKUSSION UND FAZIT</u></b>	<b>324</b>
<b><u>ZUSAMMENFASSUNG DER DISSERTATION</u></b>	<b>I</b>
<b><u>ABBILDUNGSVERZEICHNIS</u></b>	<b>II</b>
<b><u>TABELLENVERZEICHNIS</u></b>	<b>IV</b>
<b><u>ABKÜRZUNGSVERZEICHNIS</u></b>	<b>V</b>
<b><u>LITERATURVERZEICHNIS</u></b>	<b>VII</b>
<b><u>ANHANG</u></b>	<b>XXX</b>
<b><u>DANKSAGUNG</u></b>	<b>LXXIX</b>

# 1 Einleitung

*„Es mag in der Natur des Menschen liegen, Fehler zu machen, aber es liegt ebenso in der Natur des Menschen, Lösungen zu entwickeln, bessere Alternativen zu finden und sich den Herausforderungen der Zukunft zu stellen.“ (IOM-Report)*

Weltweit werden jährlich eine viertel Milliarde und in Deutschland allein über 12 Millionen Operationen durchgeführt [9, 172]. Das klinische Fachpersonal verbringt hierbei Höchstleistungen. Wesentlich für eine erfolgreiche Operation sind eine gute Planung, Durchführung und Nachbehandlung, welche den Einsatz komplexer Medizintechnik vor, während und nach der Operation erfordern. Wegen der meist invasiven Manipulation am Patienten handelt es sich bei einem Operationssaal um einen hoch sicherheitskritischen Bereich, in dem Fehler schwerwiegende Folgen haben können. Dieser multidisziplinäre Mehrpersonenarbeitsplatz ist durch leistungsbegrenzende Faktoren wie Klima, Stress, Müdigkeit, hierarchische Barrieren etc. gekennzeichnet [321]. Zusätzlich wird das medizinische Fachpersonal durch eine ständig zunehmende und komplizierter werdende Gerätemedizin belastet. Benutzungsfehler durch unzureichend gestaltete und unergonomische Medizintechnik sind keine Seltenheit. Immer mehr Zeit, welche zur Behandlung des Patienten gedacht ist, muss für die Technikbedienung aufgewendet werden. Ergonomische Defizite, gepaart mit hohem physischen und physikalischen Leistungsaufwand, zeitlichem Druck und hohem Risiko führen zu einem gesteigerten Gefahrenpotential für den Patienten und das medizinische Team [202]. Aktuelle Studien zufolge ist mangelnde Gebrauchstauglichkeit eine häufige Ursache für Fehler und die daraus resultierenden Risiken und Gefahren für Personal und Patienten im medizinischen Arbeitsfeld. Ergonomische und gebrauchstaugliche Medizingeräte sind daher zentrale Elemente für die Patientensicherheit, für effiziente Arbeitsabläufe und zufriedene Anwender. Durch gebrauchstaugliche Geräte können die Beanspruchung und Belastung des Benutzers reduziert, die Arbeitssicherheit erhöht, die Arbeitsergebnisse verbessert, die Anwenderzufriedenheit gesteigert und die Wirtschaftlichkeit vergrößert werden [48]. Die Beachtung von ergonomischen Gesichtspunkten und Gebrauchstauglichkeitsanforderungen ist daher bei der Optimierung des medizinischen Arbeitsplatzes unvermeidbar, und die Auswahl passender Medizintechnik ist für jedes Krankenhaus als erfolgsfördernd zu sehen. Doch längst kommt der Benutzungsqualität nicht die Aufmerk-

samkeit zu, wie sie funktionalen Produkthanforderungen gewidmet wird. Eine mögliche Ursache liegt darin, dass Gebrauchstauglichkeit schwer zu bewerten und zu kennzeichnen ist. In dieser Arbeit soll ein Bewertungsschema entwickelt werden, welches die Auswahl gebrauchstauglicher Medizintechnik vereinfacht. Durch eine neue Gesetzgebung und harmonisierte internationale Standards [97, 104, 132] werden den Herstellern medizinischer Produkte Aktivitäten für die Entwicklung gebrauchstauglicher Medizinprodukte auferlegt. Die internationalen Normen DIN EN IEC 60601-1-6 und DIN EN IEC 62366 fordern eine frühzeitige Verankerung von Ergonomie und Gebrauchstauglichkeit in den ersten Phasen der Entwicklung eines Medizinproduktes. Explizite Methoden zur Erfassung und Bewertung der Gebrauchstauglichkeit medizinischer Produkte sind bisher noch nicht vorhanden. Traditionelle Forschungsansätze der Mensch-Maschine-Kommunikation sind essenziell, aber nicht ausreichend für die Komplexität typischer moderner Geräte im sicherheitskritischen Kontext der Medizin [386]. Verfahren der Risikoabschätzung von Gebrauchstauglichkeitsproblemen in der Medizin existieren noch nicht. Praxistaugliche, prospektiv ausgerichtete Methoden zur gebrauchszentrierten, ergonomischen Produktgestaltung, die ergonomische Produkthanforderungen definieren und testen und damit zukünftige Produkte ergonomisch gestalten helfen, werden nicht nur durch den Fachausschuss „Ergonomie“ der Deutschen Gesellschaft für Biomedizinische Technik im VDE (DGBMT) gefordert [78, 163]. Im Rahmen dieses Projektes soll mit ausführlichen Studien eine Standardprozedur entwickelt werden, welche es ermöglicht, medizintechnische Produkte im Hinblick auf ihre Gebrauchstauglichkeit zu überprüfen und zu bewerten. Hierbei steht nicht die zu entwickelnde Technik, sondern der Mensch, der diese Technik bedienen muss, im Mittelpunkt. Für ihn muss ein einfacher Zugang zur Medizintechnik geschaffen werden. In dieser Arbeit, speziell für medizinische Produkte entwickelte Bewertungsmaßstäbe und Evaluationsmethoden sollen dies zukünftig gewährleisten.

## **1.1 Vorgehensweise und Aufbau der Arbeit**

Diese Arbeit und die in ihr vorgestellten Evaluations- und Bewertungsverfahren zur Gebrauchstauglichkeit medizinischer Geräte sollen die Gebrauchstauglichkeitsevaluation im Rahmen des in den Europäischen Richtlinien DIN EN IEC 60601-1-6 und DIN EN IEC 62366 geforderten Gebrauchstauglichkeitsent-

wicklungs-Prozesses unterstützen. Die entwickelte Methode soll ein standardisiertes Rahmenwerk liefern, mit dem die Gebrauchstauglichkeit medizinischer Geräte bestimmt werden kann.

Basierend auf dem zweigeteilten Ansatz der Gebrauchstauglichkeitsevaluation und der Gebrauchstauglichkeitsbewertung kann die Gebrauchstauglichkeit medizinischer Produkte bestimmt werden. Hierfür verknüpft die Methode Elemente der Gebrauchstauglichkeitsevaluation mit Elementen des Risikomanagements. Erst durch diese Kombination kann die Gebrauchstauglichkeit unter dem Aspekt der Sicherheit betrachtet werden. Durch die standardisierte Vorgehensweise wird die Reproduzierbarkeit der Ergebnisse im Sinne von Reliabilität und Objektivität erreicht.

Im ersten Kapitel dieser Arbeit wird auf die theoretischen Grundlagen des Themenbereiches der Mensch-Maschinen-Kommunikation im Allgemeinen und für die medizinische Arbeitsumgebung im Speziellen eingegangen. Wichtige theoretische Rahmenbedingungen für den an dieser Kommunikation beteiligten Menschen und die Benutzerschnittstelle werden erläutert, um schließlich eine Spezifikation für das medizinische Arbeitssystem, für medizinische Geräte und deren Benutzer vornehmen zu können.

Das zweite Kapitel gilt dem Themenbereich der Gebrauchstauglichkeit. Nach Vorstellung derselben und des benutzerzentrierten Entwicklungsprozesses wird auf die Notwendigkeit gebrauchstauglicher Geräte im medizinischen Arbeitssystem, auf den Bezug zum Thema dieser Arbeit und auf die Notwendigkeit einer standardisierten Methode zur Ermittlung der Gebrauchstauglichkeit für medizinische Geräte eingegangen. Relevante gesetzliche Vorschriften und Normen werden erläutert. Weiterhin werden die Methoden der Gebrauchstauglichkeitsevaluation allgemein vorgestellt und verglichen, um deren Relevanz für diese Arbeit zu diskutieren.

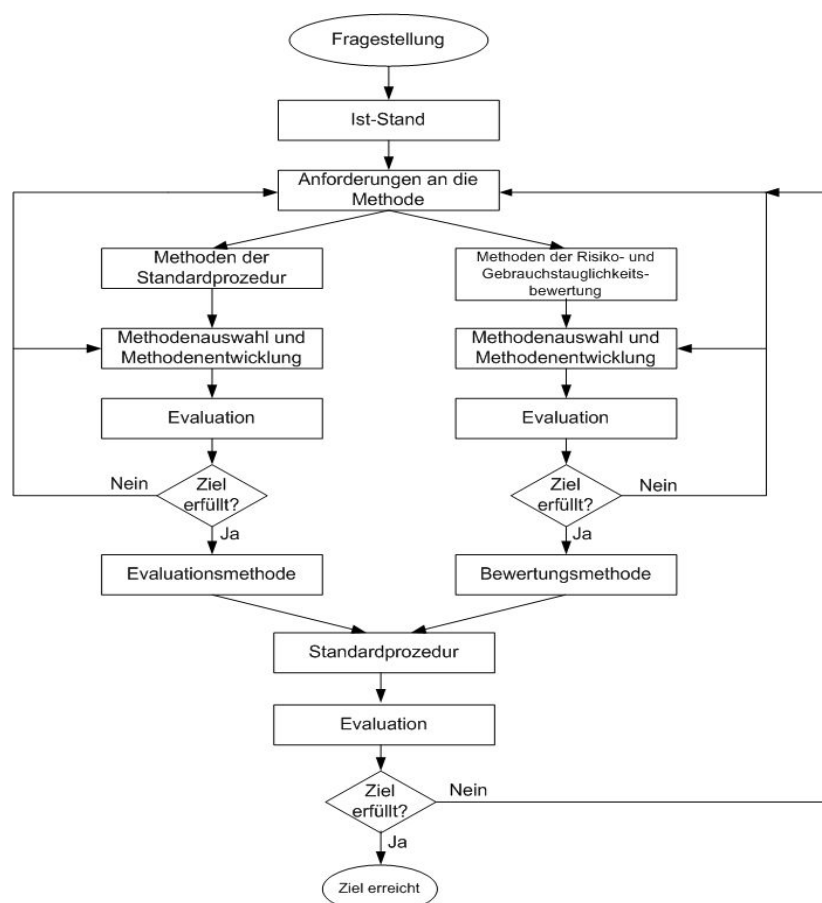
Kapitel 3 dient der Vorbereitung der Methodenentwicklung. In diesem Kapitel werden bestehende Prüfverfahren für medizinische Geräte vorgestellt und deren Bedeutung für die Standardprozedur diskutiert. Weiterhin werden Anforderungen an die Methode hergeleitet.

Die aus Evaluations- und Bewertungsteil bestehende Standardprozedur wird in zwei Phasen entwickelt und evaluiert. Kapitel 4 handelt von der Entwicklung der Evaluationsmethode. Nach theoretischen Herleitungen der Methodenelemente und entsprechenden Basisbausteinen wird in einem iterativen Prozess mit ab-

wechselnden Entwicklungs- und Evaluationsphasen die Methode entworfen. Durch vier Studien werden aufgestellte Thesen verifiziert und die Methode konkretisiert.

In den folgenden beiden Kapiteln wird der Bewertungsteil der Standardprozedur entwickelt. Als erstes werden in Kapitel 5 die theoretischen Rahmenbedingungen für die Bewertung der Gebrauchstauglichkeit medizinischer Geräte hergeleitet. Hierzu werden Bereiche der Mensch-Maschine-Interaktion und des Risikomanagements herangezogen, um in verschiedenen Bewertungsmöglichkeiten verknüpft zu werden. Wie bereits bei der Entwicklung des Evaluationsteils werden in Kapitel 6 in iterativen Entwicklungs- und Evaluationsphasen die geeigneten Methoden ausgewählt und konkretisiert. In einer abschließenden Evaluationsphase werden die beiden Elemente der Standardprozedur zusammengeführt und evaluiert. Hierbei wird besonderes Augenmerk auf die Verifikation der gestellten Anforderungen und die Überprüfung der Testgütekriterien Objektivität, Reliabilität und Validität gelegt.

In Abb. 1 ist der Entwicklungsprozess und somit auch das Konzept dieser Arbeit grafisch dargestellt.



**Abb. 1: Iterativer Entwicklungsleitfaden dieser Arbeit**



## 1.2 Theoretische Grundlagen

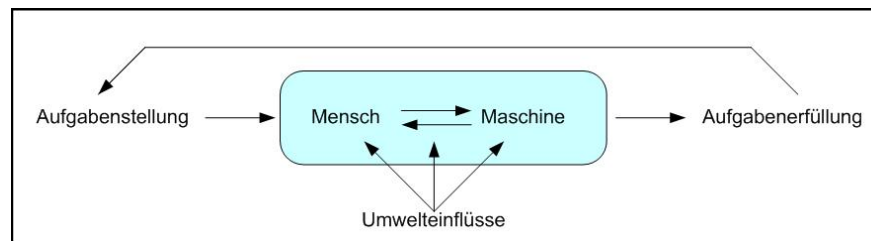
### 1.2.1 Mensch-Maschine-Kommunikation

Das in dieser Arbeit behandelte Gebiet der Mensch-Maschine-Kommunikation (MMK), teilweise auch als Mensch-Maschine-Interaktion (MMI) bezeichnet, beschäftigt sich mit dem gegenseitigen Austausch von Informationen zwischen Menschen und einem technischen Gerät über eine Schnittstelle [118]. Bullinger versteht unter einem Mensch-Maschine-System „alle Komponenten eines Arbeitssystems zur funktionellen Interaktion zwischen dem Menschen und einem technischen System“ [47]. Bei dieser Definition wird jedoch unterstellt, dass die Kommunikation zwischen dem Menschen und einer Maschine funktionell sein muss. Johannsen hingegen beschränkt sich in seiner Definition eines Mensch-Maschine-Systems auf „[...] das Zusammenwirken eines oder mehrerer Menschen mit einem technischen System“ [206]. In diesem System werden Informationen ausgetauscht, wobei die Eingaben des Nutzers im technischen Gerät Reaktionen auslösen, die wiederum an den Benutzer zurückgegeben werden. Timpe bezeichnet dies als einen gezielten Informationsaustausch, der zwischen einem Menschen und einer Maschine innerhalb von festgelegten Systemgrenzen stattfindet, um fremd- oder selbstgestellte Aufgaben zu erfüllen [391]. Der Austausch von Informationen kann dabei auf direktem oder indirektem Weg erfolgen. Die Komplexität der Maschine ist hierbei irrelevant, da der Begriff der Maschine sowohl einen einfachen Kugelschreiber als auch einen hoch automatisierten Steuerstand umfassen kann.

Bei der Interaktion zwischen Mensch und Maschine wird bidirektional eine Benutzerschnittstelle verwendet. Auf der Nutzerseite werden die Informationen über Eingabegeräte verschiedener Art realisiert. Dies können einfache Knöpfe, Steuerelemente oder eine Tastatur sein, auch eine Spracheingabe ist möglich. Die Informationsausgabe der Maschine erfolgt über diverse Anzeigen. Die direkte Interaktion erfolgt ausschließlich auf dem Weg von der Maschine zum Empfänger, wobei Umgebungsinformationen der Maschine direkt durch natürliche Parameter abgebildet werden, beispielsweise durch das „Zischen“ eines Ventils.

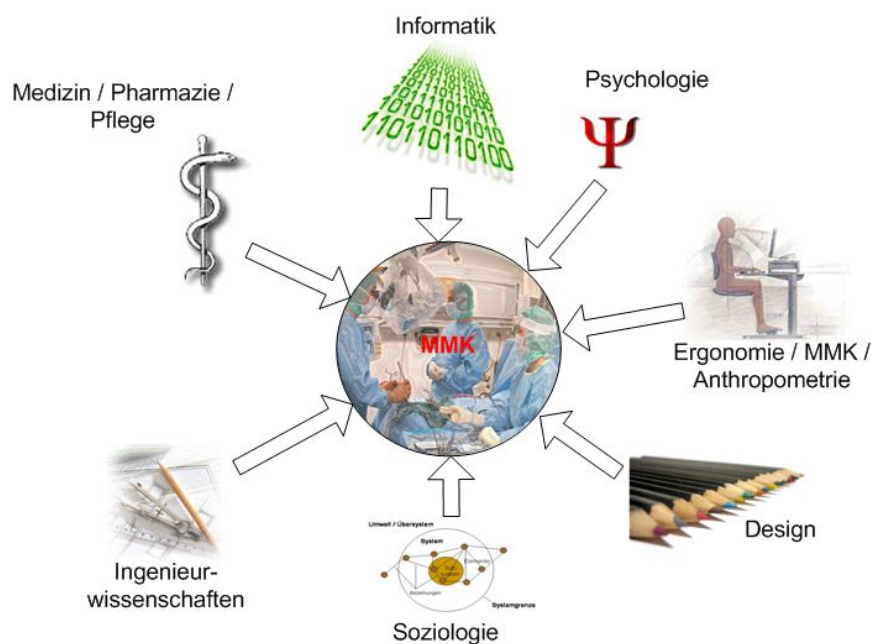
Grundsätzlich gilt für Mensch-Maschine-Systeme das allgemeine Wirkungsschema menschlicher Arbeit [135]. Auf das bestehende Wirkungsgefüge kann die Umwelt störend einwirken, indem sie beispielsweise die Fähigkeit des Menschen negativ beeinflusst (z.B. durch sein soziales Umfeld), den Informationsfluss vom

Menschen zur Maschine oder umgekehrt stört (z.B. durch Lärm) oder die Wirkung der Maschine selbst beeinflusst (z.B. durch mech. Schwingungen) (Abb. 2). Bereits Shannon & Weaver erwähnen in ihrem Kommunikationsmodell diese Störungen [367].



**Abb. 2: Allgemeines Wirkungsschema menschlicher Arbeit [135]**

Als oberste Prämisse der Forschung im Bereich der MMK gilt es das Mensch-Maschine-System zu optimieren und den Bedürfnissen des Menschen anzupassen. Hierzu sollten menschliche Belange nicht nur berücksichtigt, sondern in den Vordergrund des Interesses gerückt werden. Deshalb ist der alleinige Fokus auf informationstechnische Aspekte (d.h. auf die Maschinen) bei weitem nicht ausreichend. Psychologische Aspekte (d.h. die Menschen) sollten gleichwertig mit einbezogen werden [58]. Um diese benutzergerechte Gestaltung eines Mensch-Maschine-Systems im medizinischen Kontext durchzuführen, ist eine interdisziplinäre Zusammenarbeit der Bereiche Medizin, Ingenieurwissenschaften, Informatik, Psychologie, Arbeitswissenschaft, Kognitionswissenschaft, Ergonomie, Soziologie und Design unumgänglich (Abb. 3).



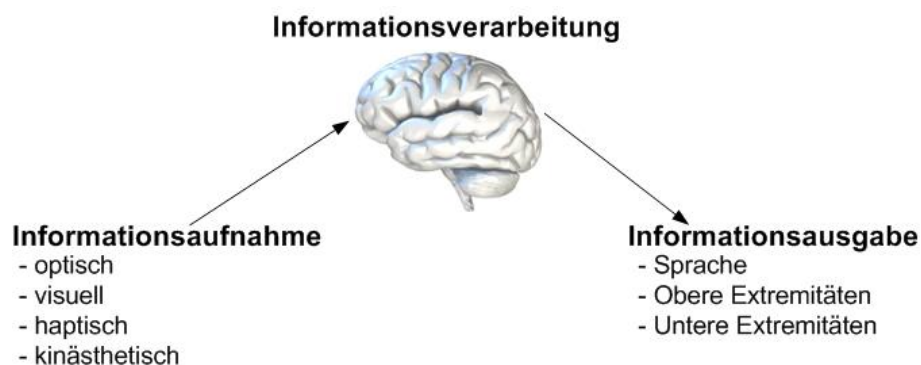
**Abb. 3: Interdisziplinarität der MMK in der Medizin**

Die Zeiten, in denen die Funktionalität der Geräte im Vordergrund stand, nur speziell geschultes Personal in der Lage war, diese zu bedienen, und nur wenig Rücksicht auf Benutzerfreundlichkeit genommen werden konnte, müssten schon lange vorbei sein. Es sollte daran gearbeitet werden, Systeme zu entwickeln, die von jedem ohne besondere Vorkenntnisse genutzt werden können. Furcht bzw. Frustration, die bei der Arbeit mit diesen Systemen entstehen können, sollten verringert werden [46, 314, 368]. Heute müsste sich der Anwender, dank technischer Weiterentwicklung und stetiger Forschung auf dem Gebiet der Mensch-Maschine-Kommunikation, nicht mehr ausschließlich nach der Maschine richten. Raskin verlangt, dass die Maschine die Bedürfnisse des Nutzers aufgreift. So sollte das User-Centered-Design<sup>1</sup> im Rahmen der Entwicklung von Kommunikationsschnittstellen die zentrale Rolle spielen [324].

### 1.2.1.1 Das Subsystem Mensch

Das wichtigste Glied in der MMK ist der Mensch. Systeme müssen sich nach dem Menschen richten und nicht der Mensch nach dem System [324]. Um gebrauchstaugliche Geräte zu entwickeln und die Gebrauchstauglichkeit dieser Geräte zu bewerten, ist das Verständnis der menschlichen Informationsaufnahme und des menschlichen Verhaltens elementare Voraussetzung.

Ebenso wie die Kombination von Mensch und Maschine als ein System mit einem Eingang (Aufgabenstellung) und einem Ausgang (Aufgabenerfüllung) verstanden wird, kann der Mensch gewissermaßen als ein Subsystem angesehen werden. Auch hier gibt es mehrere Unterelemente, welche durch ihre Ein- und Ausgänge gekennzeichnet sind. Eine grobe Unterteilung in Informationsaufnahme, Informationsverarbeitung und Informationsausgabe ist in Abb. 4 dargestellt.

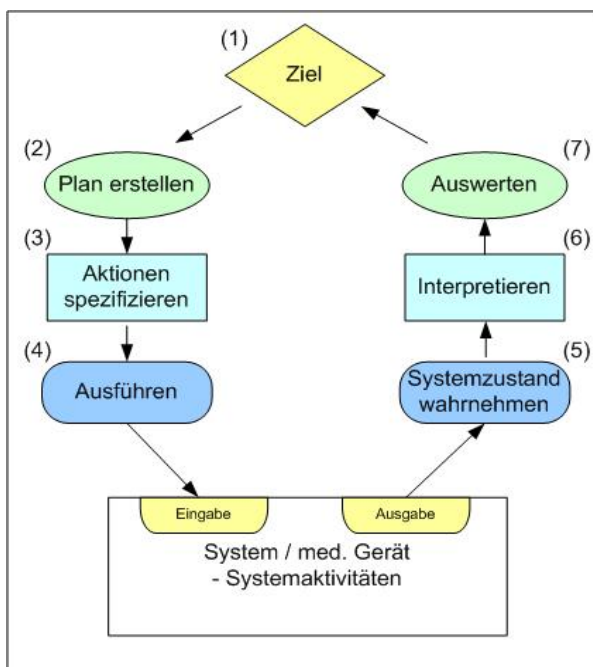


**Abb. 4: Menschliche Informationsaufnahme und -ausgabe [adaptiert von 135]**

<sup>1</sup> User-Centered-Design: Benutzerorientiertes Design (► Kap. 2.1.4)

Zur Vereinfachung des komplexen Gesamtgefüges Mensch wird dieser meist als Modell veranschaulicht. Man unterscheidet hierbei das Wahrnehmungs-, das Bewegungs- und das kognitive System. Das Wahrnehmungssystem steht am Anfang der Informationsverarbeitung. Es besteht aus verschiedenen Wahrnehmungskanälen, welche Reize und damit verbundene Informationen aufnehmen und in Form von elektrischen Impulsen über Nervenbahnen dem Gehirn als Ort der Informationsverarbeitung zuführen.

Zur weiteren Informationsverarbeitung werden die aufgenommenen Informationen dem kognitiven System zugeführt. Das Gehirn interpretiert in die wahrgenommene Information eine Bedeutung und trifft Entscheidungen. Die „neue“ Information wird zunächst temporär gespeichert und kann durch mehrfaches Bestätigen längerfristig gespeichert werden, andernfalls wird sie vergessen. Die menschliche Informationsaufnahme ist allerdings beschränkt. Grenzen und Eigenschaften für das menschliche Leistungsvermögen (Hörschwelle, Abhängigkeit von Größe und Kontrast eines Sehobjektes, Vibrationsschwelle etc.) sind sehr gut untersucht. Entsprechende Angaben lassen sich z.B. bei Eichstätt und Schmidtke [135] finden. Nach der Entscheidungsfindung wird das Bewegungssystem, bestehend aus einem passiven (Knochen, Bänder, Gelenke) und einem aktiven Anteil (Muskeln, Sehnen, Sehnenscheiden, Schleimbeutel) aktiviert, um die motorischen Funktionen des Menschen zu steuern.



**Abb. 5: Phasen der Aktionstheorie nach Norman [295]**

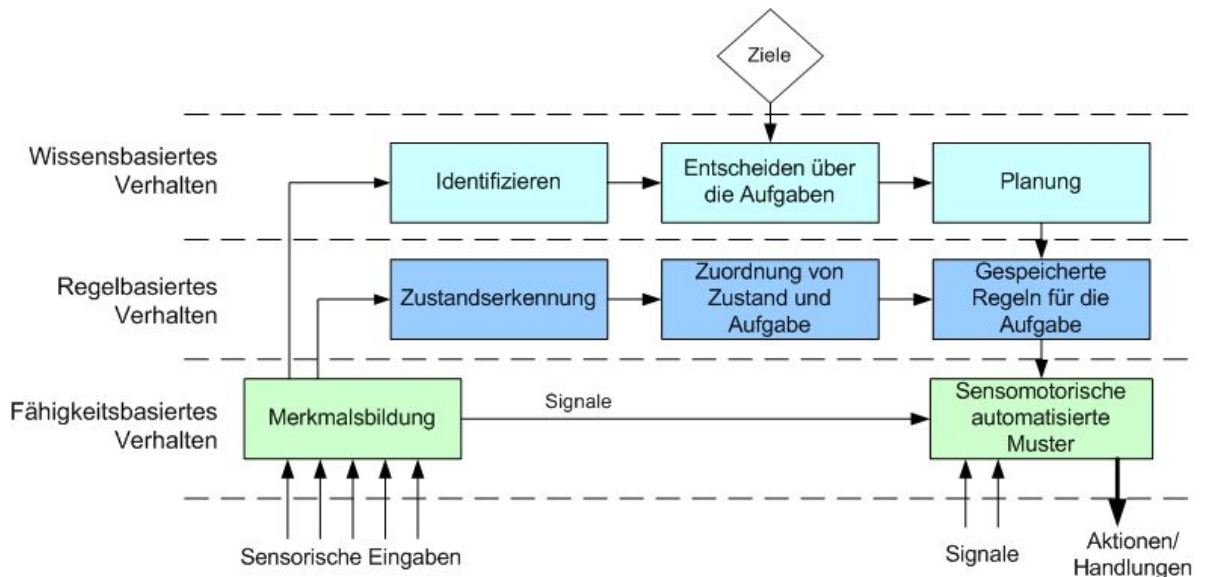
Mit Hilfe der Informationsaufnahme, der Informationsverarbeitung und des Bewegungssystems können Nutzer von Geräten Aufgaben und Arbeiten verrichten, um die von ihnen gestellten Ziele zu erreichen. Zur Veranschaulichung dieses Vorgangs kann Normans aus sieben Phasen bestehende Aktionstheorie herangezogen werden, welche für die Mensch-Computer-Interaktion und das Design von Benutzerschnittstellen erstellt wurde [127, 292, 295] (Abb. 5). Der Benutzer entwickelt zur Erreichung seines Ziels (1) einen Plan (2), in dem

er die Aktionen spezifiziert (3), mit denen er das Ziel erreichen möchte. Nach der Ausführung (4) der Aktionen wird der Systemzustand wahrgenommen (5) und ausgewertet (6). Schlussendlich werden das Ergebnis beurteilt (7) und ggf. weitere Ziele formuliert (1). Norman stellt die Forderung, dass Systeme Aktionen bereitstellen müssen, welche mit den Absichten der Nutzer übereinstimmen. Er bezeichnet die Differenz zwischen den Benutzerintentionen und den vorgesehenen Aktionen als Ausführungskluft („gulf of execution“). Deshalb sind Ziele und Aufgaben zur Zielerreichung unbedingt getrennt voneinander zu betrachten.

Benutzer arbeiten dementsprechend zielorientiert [52] und bedienen sich zur Zielerreichung einer Arbeitsaufgabe, welche der „zur Zielerreichung erforderlichen Aktivität“ entspricht [113]. Arbeitsaufgaben sind dabei nicht nur als Funktionen oder Leistungsmerkmale des Produktes anzusehen. Sie müssen vielmehr auf die angestrebten Ziele bezogen sein. Normans Sicht auf den Aktionszyklus menschlicher Handlung zeigt, dass Menschen oft mehrere Möglichkeiten haben, durch ihre Arbeitsaufgaben ihre Ziele zu erreichen. Zum Beispiel hat der Mensch für die Beschaffung eines warmen Essens die Optionen: a) den Herd oder b) die Mikrowelle zu verwenden, c) das Auto zu nehmen, um Essen zu holen oder d) zu telefonieren, um Essen zu bestellen. Anwender können somit ihre Ziele mit einer oder mehreren Arbeitsaufgaben erreichen [303]. Um komplexere Ziele zu erreichen, unterteilen sie diese in Subziele, welche wiederum Aufgaben und Teilaufgaben zur Zielerreichung erfordern [4, 83, 218, 312]. Diese können erneut unterteilt werden, bis hin zu einzelnen Handlungen (Aktionen) [160, 292, 312, 314], welche atomaren, nicht weiter unterteilbaren Nutzeraktionen entsprechen [46]. Nach Preece et al. besitzen letztere keine Problemlösungs- oder Kontrollstrukturkomponente [314].

Um diese Prozesse menschlicher Bedientätigkeit zu analysieren, kann das Leistungsmodell von Rasmussen [326] herangezogen werden. Rasmussen unterscheidet drei Verarbeitungsebenen bei der Bedientätigkeit: Die fähigkeitsbasierte Ebene bezieht sich auf bereits abgespeicherte Muster, d.h. auf bekannte Routinen. Bei der regelbasierten Ebene greift der Mensch auf bestehende und bekannte Regeln zurück, welche, wenn sie aktiviert werden, entsprechend einer Wenn-dann-Beziehung ein Verhalten auslösen. Die wissensbasierte Ebene wird bei unbekanntem Situationen angewandt [234, 326]. Bei diesen neuen Situationen muss der Mensch aktiv planen und bewusst auf analytische Prozesse und gespeichertes Wissen zurückgreifen. Er analysiert die Beziehungen zwischen Struk-

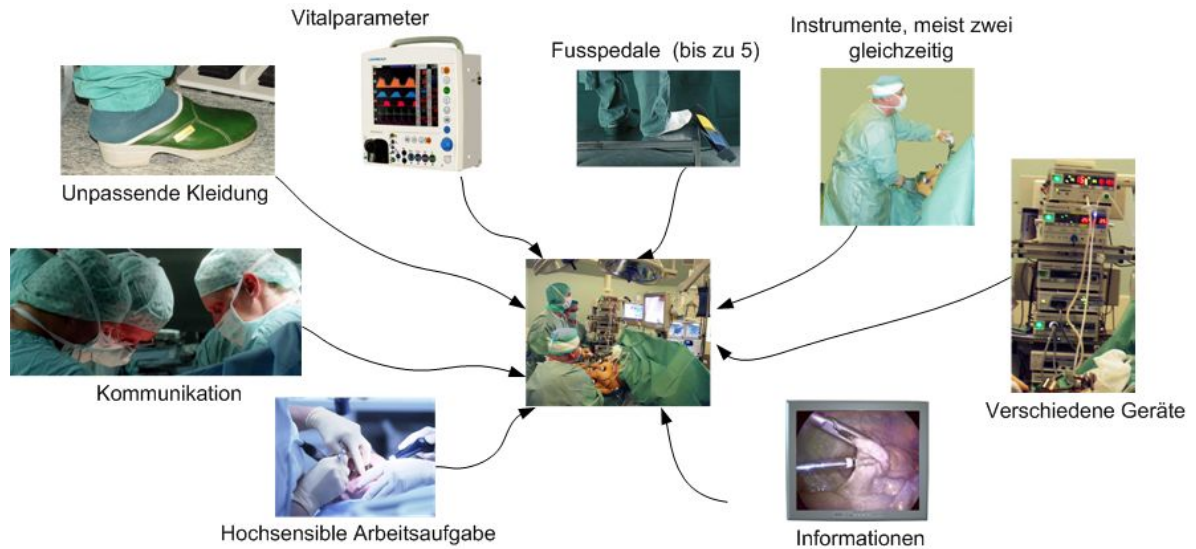
tur und Funktion, leitet eine Diagnose ab, formuliert die Korrekturhandlungen, handelt und überwacht das Ergebnis (Abb. 6).



**Abb. 6: 3-Ebenen-Modell des menschlichen Verhaltens bei Bedientätigkeiten [326]**

In der Regel werden Menschen mit unterschiedlichen Informationen gleichzeitig konfrontiert. Entscheidend bei einer gleichzeitigen Informationsaufnahme ist das Aufnahmevermögen. Wenn zu viel Information bereitgestellt wird, dann kann dies zu einem Datenüberlauf („data overload“) führen. Bekannte Grenzen für das Kurzzeitgedächtnis sind  $7 \pm 2$  Informationseinheiten (Chunks<sup>2</sup>). Das heißt, dass der Mensch nur fünf bis neun parallele Informationen gleichzeitig aufnehmen und bearbeiten kann. In vielen Arbeitssystemen sind Anwender durch eine zu hohe Zahl an Informationen überlastet. Abb. 7 veranschaulicht eine Informationsüberflutung im medizinischen Kontext. Forderungen an die Hersteller, die verfügbare Information gering zu halten und nur die essentiellen Informationen bereitzustellen, sind daher gerechtfertigt [346, 415].

<sup>2</sup> Einheit, die bewusst und aufmerksam aufgenommen und kurzfristig behalten werden kann. Ein Chunk kann ein grafisches Zeichen, ein Signal oder ein Satz sein.



**Abb. 7: Einflüsse im medizinischen Arbeitssystem am Beispiel OP**

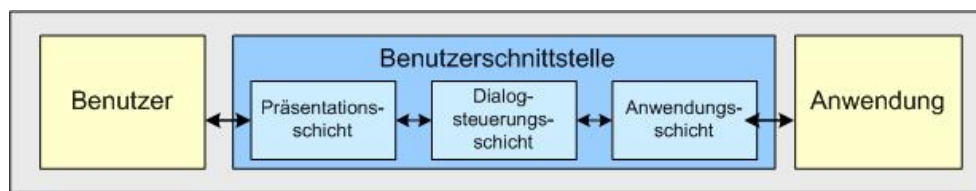
### 1.2.1.2 Benutzerschnittstellen

Ein besonderes Merkmal der MMK ist die Verwendung einer Schnittstelle (engl. Interface), die als Bindeglied zwischen Maschine und Nutzer und zwischen Nutzer und Maschine dient. Eine Benutzerschnittstelle umfasst alle Funktionen zur Kontrolle des Gerätes, wie z.B. Schalter, Knöpfe und Eingabemedien sowie alle Funktionen, welche dem Nutzer Informationen bereitstellen, wie z.B. akustische Meldungen, Anzeigen, Displays und Leuchten. Weiterhin inkludiert die Benutzerschnittstelle die Logik, wie ein System auf Eingaben des Nutzers reagiert und welche Informationen es wann bereitstellt. Die internationale Norm DIN EN IEC 60601-1-6 definiert eine Benutzer-Gerät-Schnittstelle als „Mittel, mit dem der Anwender (Benutzer) mit dem Gerät kommuniziert“ [97].

Benutzerschnittstellen ermöglichen, dass der unterschiedliche Zeichenvorrat des Menschen und der Maschine durch eine Verbindung der jeweils einen oder anderen Seite verständlich gemacht wird. Zum Beispiel verändern Bedienhandlungen nicht direkt den Zustand des Systems, sondern sie müssen erst in Kommandos transferiert werden, um für das System verständlich zu sein. Sowohl der Benutzer als auch die Geräteanwendungen tauschen Informationen mit der Benutzerschnittstelle aus und interagieren auf diese Weise miteinander. Nach Preims Definition geht eine Benutzerschnittstelle noch über diese Interaktion hinaus: „Eine Benutzerschnittstelle im weiteren Sinne besteht aus allen Interaktionsmöglichkeiten zwischen einem Menschen und einem Computersystem, einschließlich Einführungsprogrammen (Tutorials, Guided Tours), Hilfsfunktionen und Handbü-

chern. Die Benutzerschnittstelle steht als Mittler zwischen einem Anwendungsprogramm und einem Benutzer; sie macht die Funktionen der Anwendung zugänglich“ [315].

Zur Beschreibung von Benutzerschnittstellen können verschiedene Modelltypen herangezogen werden. Mit dem Seeheim-Modell sollte ein Standard für das Design von Benutzerschnittstellen festgelegt werden. Das 1983 entwickelte Modell hat sich im Rahmen des Benutzerschnittstellenmanagements etabliert. Wie in Abb. 8 dargestellt, wird eine Unterteilung der Benutzerschnittstelle in eine Präsentations-, Dialogsteuerungs- und Anwendungsschicht vorgenommen [164, 315].



**Abb. 8: Das Seeheim-Modell [315]**

Die Präsentationsschicht oder Darstellungsschicht ist der statische, sichtbare Teil der Bedienoberfläche. Diese Schicht ist der Teil der Schnittstelle, der dem Anwender visuelle und akustische Ein- und Ausgaben übermittelt. Der Benutzer agiert über die Präsentationsschicht mit der Dialogsteuerungsschicht. Diese stellt den dynamischen Teil der Bedienoberfläche dar. Sie übernimmt Anforderungen der Präsentationsschicht zur Weiterleitung an die Anwendung und umgekehrt. Im Allgemeinen dient diese Ebene der Steuerung des Informationsflusses zwischen Eingabe- und Ergebnisverarbeitung. In der Anwendungsschicht spielt sich die zugrunde liegende Anwendung ab. Sie ist für die Umsetzung der Benutzereingaben in Aktionen verantwortlich und wird durch das Zusammenwirken von Präsentationsschicht und Dialogsteuerungsschicht gelenkt [315].

Ein weiteres Modell ist das IFIP-Modell der Arbeitsgruppe International Federation for Information Processing (IFIP). Dieses Schichtenmodell, das vor allem der Bewertung von interaktiven Systemen dient, ist im Gegensatz zu dem zuvor beschriebenen Seeheim-Modell nicht von Computergrafikern und Softwaretechnikern geprägt worden, sondern von Ergonomen und Psychologen. Es berücksichtigt nicht nur die internen Komponenten einer Benutzerschnittstelle, sondern auch die (arbeits-)organisatorische Umgebung (Abb. 9).



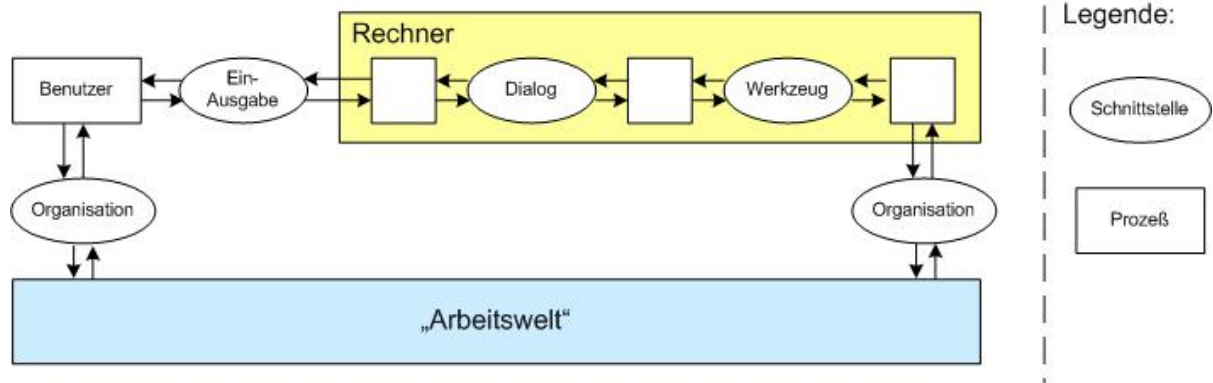


Abb. 9: Das IFIP Modell für Benutzerschnittstellen [127]

Beim IFIP-Modell wird eine Benutzerschnittstelle durch mehrere Prozesse und Schnittstellen beschrieben. Die Ein- und Ausgabeschnittstelle ist verantwortlich für die physischen Eingabetechniken und die Ausgabemöglichkeiten des Systems. Über die Dialogschnittstelle werden die Dialogsteuerung vorgenommen und die Interaktionsstile festgelegt. Die Werkzeugschnittstelle ist äquivalent zu der Anwendungsschicht des Seeheim-Modells. Anders als im Seeheim-Modell ist hier eine Organisationsschnittstelle vorhanden, welche der Eingliederung einer Benutzerschnittstelle in das Umfeld dient [127, 315]. Durch diese Einordnung des Benutzers und des Gerätes in den Kontext der Organisation hat die Arbeitsorganisation Einfluss auf den Benutzer, während die Systemorganisation Einfluss auf das Gerät ausübt.

Benutzer selbst erkennen jedoch lediglich die Benutzerschnittstelle eines Gerätes und schließen von dieser auf das komplette Gerät. Dementsprechend sind Benutzerschnittstellen die Visitenkarten eines Gerätes. Werden bei einer Entwicklung von Geräten ergonomische Aspekte für die Benutzerschnittstelle eingehalten, kann davon ausgegangen werden, dass die Akzeptanz bei den Nutzern hoch ist [189]. Im Rückkehrschluss ist anzunehmen, dass schlecht gestaltete Benutzerschnittstellen auf das ganze Gerät oder System zurückgeführt werden. In einer Vielzahl von Situationen verzweifeln Menschen am Umgang mit der Technik. Nahezu jeder kann von derartigen Situationen berichten, und auch die Literatur gibt mehr oder weniger bekannte Beispiele von kleinen Ärgernissen bis hin zum katastrophalen Interaktionsversagen mit Todesfolge [248, 370, 383].

Am 3. Juli 1988 starben 290 Passagiere an Bord des Airbus Flug 655 der Iran Air aufgrund einer tragischen Fehlinterpretation während des Iran-Irak Konflikts. Ursache war, dass ein Operator auf dem Kriegsschiff USS Vincennes das zivile Flugzeug mit einem iranischen Jagdflugzeug verwechselte. Den Operator hätte hierbei die zunehmende Höhe des vermeintlichen Angreifers warnen müssen, denn ein Kampfflugzeug wäre zunächst tiefer geflogen. Obwohl der Operator an seinem System die Höhe des Flugzeuges hätte ablesen können, wurde diese wichtige Information nicht auf dem Hauptbildschirm des Waffensystems angezeigt. Der Kreuzer schoss eine Rakete ab und 290 Zivilisten starben aufgrund eines Mensch-Maschine-Kommunikationsfehlers.

Viele international agierende Unternehmen haben die Bedeutung von gebrauchstauglichen Produkten bereits erkannt. Untersuchungen belegen, dass 70 % der Software-Entwicklungskosten für die Entwicklung einer nutzergerechten Mensch-Maschine-Schnittstelle anzusetzen sind [434]. Die meisten klein- und mittelständischen Unternehmen haben hier größere Schwierigkeiten, was sich bei ihnen in einer stärkeren Funktionalitätenflut äußert [434]. Dies kennt man auch im medizinischen Bereich.

Benutzerschnittstellen und Design sind ausschlaggebend für die richtige Bedienung eines Gerätes. Einfache und schnell verständliche Bedienmöglichkeiten, sowie Konsistenz in der Anordnung der Bedienelemente sind wichtig, um Fehler zu reduzieren. Gerade bei sicherheitskritischen Geräten kann eine fehlerhafte Bedienung zu gefährlichen Situationen führen. Verschiedene Studien belegen, dass ausgerechnet Benutzerschnittstellen medizinischer Geräte unzulänglich gestaltet sind und nicht nur den Anwender bei seiner Aufgabenerfüllung schlecht unterstützen, sondern häufig sogar die Ursache für gravierende Fehler sind [17, 42, 43, 212, 220, 273, 408]. Vermeidbare Verletzungen und sogar Todesfälle im OP sind oft Folgen einer falschen Verwendung medizinischer Geräte. Menschliches Versagen ist in einigen Fällen der Grund, doch häufig liegt die Ursache in einem Problem der MMK. Nach Meinung der amerikanischen Arzneimittelzulassungsbehörde FDA (Food and Drug Administration), welche u.a. für die Sicherheit und Wirksamkeit von Medizinprodukten zuständig ist, unterstützt eine gut gestaltete Benutzerschnittstelle richtige Handlungen und vermeidet Fehlhandlungen. Informationen, Anzeigen und Kontrollfunktionen müssen hierbei im Einklang mit den Fähigkeiten, Erwartungen und Verhaltensweisen der Nutzer stehen [212]. Die Entwicklung gebrauchstauglicher Benutzerschnittstellen erfordert von den Herstellern eine genaue Beachtung gängiger Standards und Richtlinien, die Einbezie-

hung der Nutzer in den Entwicklungsprozess und eine umfangreiche Gebrauchstauglichkeitsevaluation. Der genaue Prozess wird in Kapitel 2.1.4 beschrieben.

## 1.2.2 Mensch-Maschine-Kommunikation in der Medizin

Die Mensch-Maschine-Kommunikation hat längst die Grenzen der herkömmlichen Softwareentwicklung durchbrochen. Schnittstellen kommen in fast allen Bereichen, in denen Menschen mit Geräten arbeiten, vor. Neben der Anlagensteuerung, dem privaten Unterhaltungs- und Haushaltsbereich, der Büroumgebung und der Automobilindustrie ist auch die Medizintechnik von zunehmender Mensch-Maschine-Kommunikation geprägt. Im Folgenden werden das medizinische Arbeitssystem und die beiden wichtigsten Komponenten „Geräte“ und „Anwender“ beschrieben.

### 1.2.2.1 Das medizinische Arbeitssystem

Ein Arbeitssystem ist ein „System, welches das Zusammenwirken eines einzelnen oder mehrerer Arbeitender/Benutzer mit den Arbeitsmitteln umfasst, um die Funktion des Systems innerhalb des Arbeitsraumes und der Arbeitsumgebung unter den durch die Arbeitsaufgaben vorgegebenen Bedingungen zu erfüllen“ [109]. Arbeitssysteme dienen somit der Erfüllung von Arbeitsaufgaben. Hierbei wirken Menschen und Betriebsmittel unter Umwelteinflüssen zusammen [333]. Man unterscheidet technische Systeme, welche eine Verknüpfung zwischen Maschinen betrachten, soziale Systeme, welche die Beziehungen zwischen Menschen beschreiben und soziotechnische Systeme, welche durch Beziehungen zwischen Menschen und Maschinen gekennzeichnet sind. Das medizinische Arbeitssystem kann, wie andere Arbeitssysteme auch, durch sieben Elemente beschrieben werden [333]:

1. Die **Arbeitsaufgabe**: Sie kennzeichnet den Zweck des Arbeitssystems, das Ausführen von Tätigkeiten, welche der Zielerreichung dienen.
2. Der **Arbeitsablauf**: Er beinhaltet räumliches und zeitliches Zusammenwirken von Menschen und Arbeitsmitteln, wobei die Eingabe gemäß der Arbeitsaufgabe in die Ausgabe überführt wird.
3. Die **Eingabe** eines Arbeitssystems besteht im Allgemeinen aus Arbeitsgegenständen, aber auch aus Menschen, Informationen und Energie, die im Sinne der Arbeitsaufgabe verändert oder verwendet werden.

4. Die **Ausgabe** eines Arbeitssystems besteht im Allgemeinen aus Arbeitsgegenständen, aber auch aus Menschen und Informationen, die im Sinne der Arbeitsaufgabe verändert, verwendet oder neu erstellt werden.

5./6. **Mensch und Betriebsmittel** sind Systemelemente, die gemäß der Arbeitsaufgabe die Eingabe in die Ausgabe verändern.

7. Die **Umwelteinflüsse** unterteilt man in physikalische, chemische, biologische, organisatorische und soziale Umgebungseinflüsse, die das Arbeitssystem beeinflussen.

Eine Besonderheit des medizinischen Arbeitssystems ist, dass es sich hierbei nicht um einen Arbeitsgegenstand, sondern ausschließlich um einen Patienten handelt (Abb. 10).

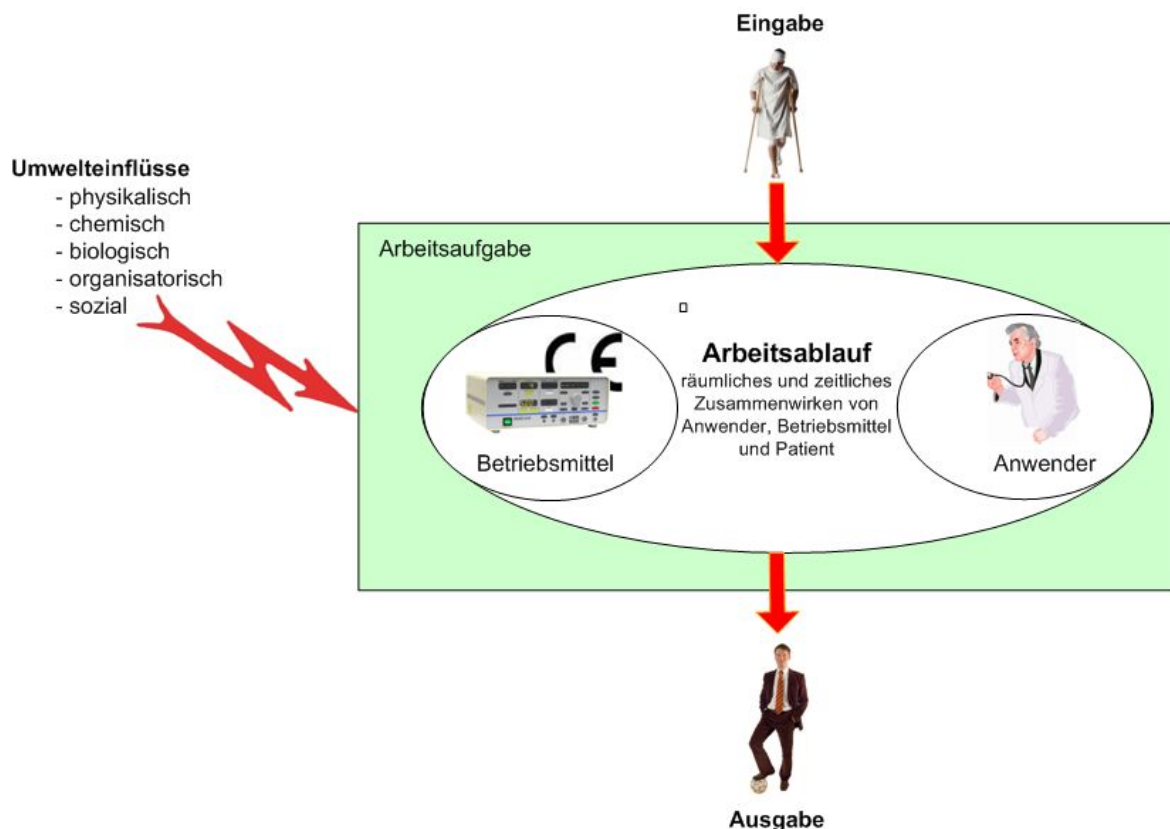


Abb. 10: Das medizinische Arbeitssystem [adaptiert von 333]

Das Ziel, mit anderen Worten, die Ausgabe des medizinischen Arbeitssystems, ist ein behandelter und zufriedener Patient. Um dieses Ziel zu erreichen, sollte innerhalb des Arbeitssystems keine zusätzliche Belastung auf die Akteure wirken, besonders nicht durch die Technik. Dieser Grundsatz wird umso bedeutender, je komplexer die Aufgaben und die daraus resultierenden Anforderungen sind.

Im Gegensatz zu den meisten anderen Arbeitssystemen ist das medizinische sehr komplex und vielfältig. Es gibt kaum ein Arbeitssystem, welches so unterschiedliche Faktoren und Variablen aufweist. Das medizinische Arbeitssystem ist geprägt durch unterschiedliche Anwender, Aufgaben, Betriebsmittel, Anwendungsorte und Arbeitsabläufe. Seine konkrete Erfassung ist kaum möglich, denn neben den stationären medizinischen Arbeitssystemen, z.B. dem OP, der Intensivstation oder der Radiologie bildet z.B. auch eine Privatperson, welche zu Hause mit Hilfe eines Blutzuckermessgerätes den Glucosespiegel misst, ebenso ein medizinisches Arbeitssystem wie ein Rettungssanitäter der Bergwacht, der bei  $-10^{\circ}\text{C}$  in 2000 Metern Höhe im Schneesturm einen verunglückten Skifahrer reanimiert.

In einem Krankenhaus stellen die medizinischen Arbeitssysteme „Operationsaal“, „Intensivstation“ und „interventionelle Radiologie“ die Kernstücke des medizinischen Arbeitssystems „Krankenhaus“ dar und sind zumeist auch die kostenintensivsten und sicherheitskritischsten Arbeitsplätze [16, 57]. Dieses System, welches sich nach nie vollständig beschreibbaren "Arbeitsgegenständen", nämlich den Patienten, ausrichtet, ist bestimmt durch die Qualität des Zusammenspiels eines Nutzers mit der Technik während einer Aufgabe in einer Arbeitsumgebung. Es ist derzeit geprägt durch einen hohen Grad technischer und wissenschaftlicher Dynamik [57]. Neue Technologien finden verstärkt Einsatz. Nutzungssituationen sind meist unterschiedlich und erfordern eine starke Individualität der medizinischen Behandlung, welche wiederum eine flexible Gestaltung dieses Arbeitsplatzes bedingt [11, 254]. Dennoch ist dieses Arbeitssystem durch fehlende Kompatibilität und Einzelkomponenten verschiedener Hersteller (Insellösungen) geprägt. Weitere physische, soziale und organisatorische Faktoren, welche dieses Arbeitssystem in Abhängigkeit von der medizinischen Behandlung und der Arbeitsumgebung mehr oder weniger stark beeinflussen, sind [2, 16]:

- mögliche Konflikte zwischen der Sicherheit der Mitarbeiter und der Patienten,
- hohe Variabilität der Aufgabenstellung, der Behandlungsmethoden, des Behandlungsverlaufs und der Verantwortlichkeiten,
- ständiger Wechsel in der Zusammensetzung des Behandlungsteams,
- hohes Stresspotenzial durch prioritären Handlungsdruck am Patienten, ständig neue Verfahrenswahl und Risikoabwägung, sowie unterschiedliche Eingriffsdauer und Arbeitszeiten,

- Einsatz hochkomplexer Geräte und Instrumente mit teilweise erheblichem Gefährdungspotenzial, unterschiedlichen Bedienkonzepten und mangelnder Kompatibilität,
- hoher Kostendruck,
- räumliche Enge am Patienten,
- Fürsorgepflicht für Personal und Patienten,
- Vertraulichkeit und Datensicherheit,
- Verantwortung,
- unterschiedliche physikalische Faktoren wie z.B. Klima, Lärm, Vibration, Beleuchtung,
- soziale Konflikte.

Jeder dieser Faktoren wirkt auf den anderen ein. Sobald einer dieser Aspekte nicht richtig „funktioniert“, entstehen Konsequenzen, in erster Linie zum Nachteil des Patienten. Unzulängliche Schnittstellen zwischen den einzelnen Faktoren sind oft die Ursache für Funktionsuntüchtigkeit.

Patientensicherheit ist somit nicht alleine das Verdienst eines sorgfältigen Anwenders, und das Auftreten von Komplikationen kann nicht automatisch mit mangelnder Sorgfalt oder Qualifikation gleichgesetzt werden. Nach Ollenschläger wird Patientensicherheit vom gesamten System der Krankenversorgung beeinflusst. Hierzu gehören u.a. die daran beteiligten Personen (Betreuungsteam), technische Bedingungen (z.B. Arbeitsplatzgestaltung), organisatorische Elemente (Organisationsroutinen, Arbeitszeiten etc.), sowie die gesellschaftlichen Rahmenbedingungen (etwa der finanzielle Rahmen der Krankenversorgung) [299].

Weitere beeinflussende Faktoren dieses Arbeitssystems sind ökonomische Zwänge. Auf der einen Seite werden die medizinischen Arbeitsaufgaben und die technische Ausstattung immer komplexer und komplizierter, auf der anderen Seite wächst der Kosten- und Zeitdruck im Gesundheitswesen immer stärker an. Gerade die Einführung der diagnosebezogenen Fallgruppen (DRG = Diagnosis Related Groups)<sup>3</sup> [219] sowie intensive Rationalisierungen und bestehender Ärzte-

---

<sup>3</sup> DRG: Diagnosis Related Groups, Diagnosebezogene Fallgruppen: Patientenklassifikationssystem, das in einer klinisch nachvollziehbaren Weise Art und Anzahl der behandelten Krankenhausesfälle in Bezug zum Ressourcenverbrauch des Krankenhauses setzt. Behandlungen werden durch Pauschalen abgegolten und selbst verursachte Mehraufwendungen werden hierdurch nicht mehr vergütet.

mangel ziehen den Wirtschaftlichkeitsrahmen enger und machen eine genaue Budgetierung notwendig.

Kostendruck und reduzierte Budgets, technologischer Fortschritt, neue wissenschaftliche Erkenntnisse sowie der demografische Wandel bringen es mit sich, dass immer weniger Personal immer mehr und immer ältere Patienten mit komplizierteren Techniken und Technologien operieren muss [337]. Die Arbeitsbelastung für das medizinische Personal nimmt im medizinischen Arbeitssystem weiter zu [16, 398]. Forderungen, wie z.B. die von Matern, nach einer effizienteren Ausnutzung der Arbeitskraft des Personals, um die Gesamtleistung des Behandlungsteams und die Sicherheit für den Patienten zu steigern [127], sind die logische Konsequenz. Dies kann jedoch nur erreicht werden, wenn dem Personal entsprechende Ressourcen in Form von einfach zu erlernenden, einfach zu verwendenden und dabei sicheren Arbeitsmitteln zur Verfügung gestellt werden. Zwar verschieben sich kritische Zwischenfälle mit zunehmender Gerätesicherheit auf menschliche Fehler, jedoch sind diese oft Folgen ungünstig gestalteter Arbeitsplätze, schlechter Arbeitsbedingungen und mangelbehafteter Arbeitsmittel [178, 363, 383]. Deshalb sollte neben der Zielverfolgung, immer bessere und schnellere Behandlungsmethoden und Arbeitsmittel zu entwickeln, auch darauf geachtet werden, dass im medizinischen Arbeitssystem möglichst keine zusätzlichen Belastungen für die Akteure durch Anhäufung komplizierter Technik entstehen. Nutzer sollten nicht durch die Geräte gezwungen werden, einen Großteil ihrer Aufmerksamkeit in deren Bedienung zu stecken. Vielmehr sollten sie ausreichend Zeit haben, sich um ihre Patienten zu kümmern. Erreicht werden kann dies durch ein möglichst optimales Zusammenspiel der Anwender und der Geräte im Arbeitssystem, durch ergonomische Gestaltung, benutzerfreundliche Bedienkonzepte und optimale Systemintegration [11, 212, 223, 259, 284, 358].

#### **1.2.2.2 Die medizinische Arbeitsaufgabe**

Entsprechend der sozio-technischen Systemtheorie [399] können medizinische Arbeitsaufgaben nach primären und sekundären Aufgaben unterschieden werden. Primäraufgaben sind zur Erfüllung der Nutzerziele erforderlich, für die das Gerät oder das System geschaffen wurde. Hierunter fallen im Sinne der Bedienung medizinischer Geräte die Aufgaben, die essentiell für die Patientenbehandlung notwendig sind. Sekundäraufgaben unterstützen die Ausführung der Primäraufgabe. Unterstützungsbereiche können die Systemerhaltung, Regulation oder

Optimierung sein [7]. Im Sinne medizinischer Arbeitssysteme sind dies Aufgaben, die nicht mit der eigentlichen Behandlung in Verbindung stehen, sondern dem Anwender bei der Erfüllung seiner Primäraufgaben Hilfestellungen gewähren. Diese Einteilung ist äquivalent der in DIN EN ISO 13407 verwendeten Einteilung in Primärziele und Sekundärziele [105].

Abstufungen in Primäraufgaben und Sekundäraufgaben sind untrennbar verknüpft mit Wahrscheinlichkeiten für einen Schadenseintritt und dem Schadensausmaß. Beispielsweise ist bei einer Fehlbedienung innerhalb einer Sekundäraufgabe (z.B. Einstellen des Kontrastes eines Anzeigeegerätes) ein Schadenseintritt und ein entsprechendes Schadensausmaß nicht zu erwarten. Anders bei einer Primäraufgabe: Wird der Bolus bei einer automatischen Spritzenpumpe falsch eingestellt, besteht eine relativ hohe Wahrscheinlichkeit für einen Schadenseintritt und ein hohes Schadensausmaß. Aus diesem Grund sind sicherheitsrelevante Aufgaben (wie z.B. „Notaus“) den Primäraufgaben zuzuordnen (Tab. 1).

	<b>Primäraufgaben</b>	<b>Sekundäraufgaben</b>
<b>Medizinische Aufgaben</b>	Aufgaben, die mit der eigentlichen Behandlung am Patienten zu tun haben	Unterstützende Aufgaben
	z.B.: Beatmen, Operieren, Gerät in betriebsbereiten Zustand versetzen, Leistung auswählen	z.B.: Helligkeit verstellen, Programme wählen
<b>Schadenseintrittswahrscheinlichkeit</b>	hoch	gering - mittel
<b>Schadensausmaßwahrscheinlichkeit</b>	mittel - hoch	gering - mittel

**Tab. 1: Medizinische Primär- und Sekundäraufgaben**

### 1.2.2.3 Medizinische Geräte

Ein medizinisches Gerät stellt eine spezielle Form eines Medizinproduktes dar. Es ist vornehmlich für medizinische Zwecke bestimmt und erreicht seine Hauptwirkung im oder am menschlichen Körper. Medizinprodukte werden in § 3 Abs. 1 [276] Medizinproduktegesetz (MPG) definiert als:



*„Alle einzeln oder miteinander verbunden verwendeten Instrumente, Apparate, Vorrichtungen, Stoffe oder andere Gegenstände einschließlich der für ein einwandfreies Funktionieren des Medizinproduktes eingesetzten Software, die vom Hersteller zur Anwendung für Menschen mittels ihrer Funktion zum Zwecke:*

- a) der Erkennung, Überwachung, Behandlung oder Linderung...von Krankheiten*
- b) der Erkennung, Überwachung, Behandlung oder Kompensierung... von Verletzungen oder Behinderungen*
- c) der Untersuchung, Ersetzung oder Veränderung des anatomischen Aufbaus oder eines physiologischen Vorgangs*
- d) der Empfängnisregelung*

*zu dienen bestimmt sind und deren bestimmungsgemäße Hauptwirkung...weder durch pharmakologische oder immunologisch wirkende Mittel noch durch Metabolismus erreicht wird...“.*

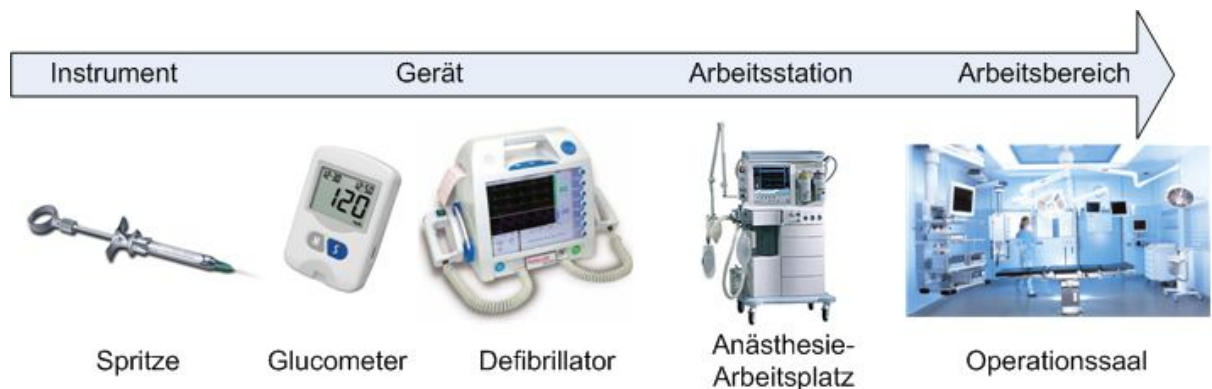
Eine durch den Ausschluss pharmakologischer, immunologischer oder metabolischer Mittel einschränkende und für ein medizinisches Gerät eindeutige Definition liefert die Richtlinie über medizinische Geräte 93/42/EWG [132]. Demnach sind medizinische Geräte:

*"Alle Instrumente, Geräte, Vorrichtungen, Materialien oder sonstigen Artikel, ob allein oder in Kombination verwendet (einschließlich der Software, die für deren vom Hersteller beabsichtigte sachgemäße Anwendung erforderlich ist), welche für Menschen eingesetzt werden sollen zum Zweck:*

- der Diagnose, Prophylaxe, Überwachung, Behandlung oder Linderung einer Krankheit,*
- der Diagnose, Überwachung, Behandlung, Linderung oder Kompensation einer Verletzung oder Behinderung,*
- der Untersuchung, des Ersatzes oder der Änderung der Anatomie bzw. eines physiologischen Prozesses,*
- der Empfängnisverhütung,*

*und die ihre gewollte Hauptwirkung im oder am menschlichen Körper nicht durch pharmakologische, immunologische oder metabolische Mittel erreichen, in ihrer Funktion jedoch von derartigen Mitteln unterstützt werden können...".*

Im Unterschied zu den meisten anderen Geräten sind medizinische Geräte sicherheitskritische Systeme, welche durch ihre vom Hersteller definierte Zweckbestimmung [121, 276] im Sinne des MPG eingesetzt werden. Demnach wird der beabsichtigte Gebrauch vom Hersteller und nicht vom Anwender festgelegt. Daraus ergibt sich, dass sich Anwender von Medizinprodukten darauf verlassen können, dass innerhalb der angegebenen Zweckbestimmung die Verwendung des Medizinproduktes in ausreichender, vom Hersteller gewährleisteter Sicherheit erfolgen kann [121]. Als Anforderung an die Anwender gilt eine nach §2 Satz 2 der Medizinprodukte-Betreiberverordnung geforderte Ausbildung oder Kenntnis und Erfahrung zum Errichten, Betreiben und Anwenden von Medizinprodukten [275].



**Abb. 11: Verschiedene medizinische Geräte bis hin zum OP-System**

Es gibt eine Vielzahl verschiedener Geräte in der Medizin, welche von einem einfachen Instrument bis hin zu komplexen Systemen reichen (Abb. 11). Grundsätzlich werden medizinische Produkte und somit auch medizinische Geräte in aktive und passive (auch „nicht aktiv“ genannte) Geräte unterteilt. Ein aktives medizinisches Gerät ist jedes medizinische Gerät, dessen Betrieb auf eine elektrische Energiequelle oder eine andere Energiequelle als die unmittelbar durch den menschlichen Körper oder die Schwerkraft erzeugte Energie angewiesen ist [131]. Somit wird weder ein Produkt, welches als reiner Überträger von Energie, Stoffen oder Parametern ohne deren Veränderung dient, noch ein sich der Energie des menschlichen Körpers bedienendes, als medizinisches Produkt angesehen. Mögliche Energiequellen sind neben Elektrizität beispielsweise: Druckluft, Vakuum, mechanische Energie einer Feder oder radioaktive Quellen. Zu den passiven Medizinprodukten gehören alle anderen, wie beispielsweise Instrumente, Optiken, Nahtmaterial und Verbandstoffe. Weiterhin können noch implantierbare und sterile Medizinprodukte unterschieden werden [106]. Ein Spezialfall sind

medizinisch elektrische Geräte, die der Normreihe DIN EN IEC 60601-1 zugeordnet werden. Diese Geräte haben oftmals einen hohen Grad an Interaktivität. Gebrauchsfehler, verursacht durch mangelnde Gebrauchstauglichkeit, sind bei ihnen häufig und geben daher zunehmend Anlass zur Besorgnis. Die Zusatznorm 6 dieser Normreihe beschäftigt sich speziell mit der Gebrauchstauglichkeit medizinisch elektrischer Geräte [173].

Oftmals bringen die zum Wohle des Menschen entwickelten medizinischen Geräte durch ihren unmittelbaren, teils invasiven Einsatz am Patienten unter möglicher Verwendung sicherheitskritischer Technik ein Sicherheitsrisiko für Patienten und Anwender mit sich, das im schlimmsten Fall bis zum Exitus führen kann [396]. Deshalb werden vom Gesetzgeber besondere Bestimmungen für Medizinprodukte festgelegt, welche für einen hohen technischen Standard sorgen sowie den Anwender und den Patienten gleichermaßen schützen. Zulassung, Verkehr und Anwendung von medizinischen Produkten werden im Zuge der Umsetzung europäischer Rechtsvorschriften in nationales Recht seit 1995 in Deutschland durch das MPG geregelt [276]. Das MPG löste die bis dahin bestehende Medizingeräteverordnung (MedGV) ab. Es verlangt vom Hersteller einen Nachweis über die medizinische Leistungsfähigkeit. Mit Hilfe von klinischen Bewertungen und Prüfverfahren wird die Freiheit von unvermeidbaren Neben- und Wechselwirkungen mit anderen Medizinprodukten, Gegenständen oder Arzneimitteln für den Patienten sichergestellt.

Als Voraussetzungen für den Vertrieb und die Anwendung als technisches Produkt in Deutschland wurden durch das MPG die Mindestanforderungen gem. § 7 und die CE-Kennzeichnung (Communauté Européenne) gesetzt, durch die der Hersteller die Einhaltung eines Konformitätsbewertungsverfahrens nachweist und nach außen signalisiert. Die Konformitätsprüfung ist vom jeweiligen Gefährdungspotential des medizinischen Gerätes abhängig [132, 276, 360]. Zu diesem Zweck werden die medizinischen Geräte in vier Klassen (I, IIa, IIb und III) eingeteilt (Tab. 2). Darüber hinaus gibt es noch die Unterklassen I<sub>s</sub> (sterile Klasse I Produkte) und I<sub>m</sub> (Klasse I Produkte mit Messfunktion). Die Klassifizierungsregeln basieren auf der Verletzbarkeit des menschlichen Körpers und berücksichtigen die potentiellen Risiken im Zusammenhang mit der technischen Auslegung der Produkte und ihrer Herstellung. Sie sind detailliert im Anhang IX der EU-Richtlinie 93/42/EWG festgelegt. Die Anwendung der Klassifizierungsregeln richtet sich nach der Zweckbestimmung der Produkte und liegt daher in der Verantwortung des Herstellers. Dies

wird auch dadurch nicht geändert, dass der Hersteller bei den Klassen IIa, IIb, und III eine benannte Stelle<sup>4</sup> in das Konformitätsverfahren mit einbeziehen muss. Je nach Klasse müssen unterschiedliche Verfahren zur Konformitätsbewertung angewandt werden [18, 31, 276]. Bei einigen von ihnen ist ein Qualitätsmanagementsystem gefordert, welches durch die Norm DIN EN 13485 definiert wird [106]. Zur Erleichterung der Klassifizierung stellt die Richtlinie 18 Klassifizierungsregeln bereit, welche in EU Leitlinien veröffentlicht wurden [262]. Grundsätzlich gilt:

- Produkte der Klasse I, mit Ausnahme solcher, die in sterilem Zustand in den Verkehr gebracht werden, und solcher mit Messfunktion, werden vom Hersteller selbst bewertet. Es ist kein Qualitätsmanagementsystem erforderlich.
- Bei Produkten der Klasse IIa ist eine benannte Stelle im Herstellungsstadium zu beteiligen.
- Produkte der Klasse IIb und III unterliegen einer Kontrolle durch eine benannte Stelle in Bezug auf die Auslegung der Produkte und die Herstellung.

Klasse I	Klasse IIa	Klasse IIb	Klasse III
Elektroden, oberflächliche Fixierhilfen, Pflaster, Gehhilfen, Mehrfachpflaster, Prothesen, Verbandmittel, Verbandstoffe	Steriles chirurgisches Nahtmaterial (Anwendung kurzfristig), Gebissimplantate, Einmalkatheter, Einmalspritzen	Anästhesiegerät, PACS, Blutbeutel, Implantate, Zahnprothesen, implantierte Prothesen	Herzkatheter, chirurgisches Nahtmaterial (Anwendung langfristig), aktive Implantate, Herzschrittmacher, Katheter, Pflaster mit Wirkstoff
Risikopotenzial: Gering	Risikopotenzial: Mittel	Risikopotenzial: Erhöht	Risikopotenzial: Hoch

**Tab. 2: Risikoklassen von Medizinprodukten**

Hersteller von Medizinprodukten dürfen diese nur dann in den Verkehr bringen, wenn kein begründeter Verdacht besteht, „dass sie die Sicherheit und die Gesundheit der Patienten, der Anwender oder Dritter bei sachgemäßer Anwendung, Instandhaltung und ihrer Zweckbestimmung entsprechender Verwendung über ein nach den Erkenntnissen der medizinischen Wissenschaften vertretbares Maß hinausgehend gefährden“ [186, 276]. Hierbei nennt der Gesetzgeber ausdrücklich die Übereinstimmung mit harmonisierenden Normen [186, 276]. Medizinische

<sup>4</sup> Benannte Stellen (engl. Notified Bodies) sind neutrale, unabhängige und kompetente Stellen, die die Konformitätsbewertung von Produkten des freien Warenverkehrs (Harmonisierungsrichtlinien Art. 95 des EG-Vertrages) durchführen, sofern dies für das jeweilige Produkt gemäß den EG-Richtlinien vorgesehen ist.

Geräte sind nur dann wertvoll, wenn sie die Aufgabe erfüllen, die Pflege der Patienten besser, einfacher, effizienter und befriedigender zu machen [408]. Sie werden häufig in unterschiedlichen Umgebungen durch verschiedene Personen angewandt, welche oft durch einen oder mehrere Faktoren beeinflusst werden, wie z.B. durch Lärm, schlechte Beleuchtung, blendende Oberflächen, Hitze, Schmutz, unsachgemäße Reinigung, elektrische Interferenzen, Feuchtigkeit, Stress und Müdigkeit [352]. Aufgrund der Vielfalt der im medizinischen Arbeitssystem benutzten Geräte können diese nur sicher und effektiv verwendet werden, wenn die Wechselwirkungen zwischen den Geräten, der Arbeitsumgebung und der im Folgenden beschriebenen Benutzer bei der Entwicklung berücksichtigt wurden. Spezielle Anforderungen an Geräte und deren Wechselbeziehung für die Ergonomie und Interaktion sind in der DIN EN 614-1<sup>5</sup> aufgestellt und können ohne weiteres auf medizinische Geräte übertragen werden [99].

#### **1.2.2.4 Benutzer medizinischer Geräte**

Ein wichtiger Teil des medizinischen Arbeitssystems sind die Benutzer (auch Anwender) medizinischer Geräte. In Anlehnung an die DIN EN 9241-10<sup>6</sup> ist ein Benutzer ein „[...]Mensch, der mit dem Produkt arbeitet“ [113]. Wie im vorherigen Kapitel bereits beschrieben, kann prinzipiell jeder Mensch potentieller Benutzer eines medizinischen Gerätes sein. Auch wenn im MPG der Benutzer nicht eindeutig definiert ist, ergibt sich aus dem Zusammenhang, dass jede Person, die am oder mit einem Medizinprodukt eigenverantwortlich Tätigkeiten ausführt, als Benutzer anzusehen ist [139]. Medizinische Geräte können dementsprechend von einer Vielzahl unterschiedlicher Nutzern verwendet werden, von gut ausgebildeten Fachärzten und Pflegepersonal bis hin zum Laien [408].

Eine genauere Abgrenzung kann erst durch die Geräte selbst, ihren Nutzungskontext und ihre Zweckbestimmung durchgeführt werden. Nutzergruppen medizinischer Geräte differieren im hohen Maße in mentalen, physischen und demografischen Eigenschaften. Es kann grob zwischen medizinischem, nichtmedizinischem Personal und Laien (Dritten) unterschieden werden. Diese Gruppen differenzieren sich weiter durch Erfahrung und Kenntnisstand (Abb. 12).

---

<sup>5</sup> DIN EN 614-1: Sicherheit von Maschinen - Ergonomische Gestaltungsgrundsätze - Teil 1: Begriffe und allgemeine Leitsätze

<sup>6</sup> DIN EN 9241 (2002): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten – Teil 10: Grundsätze der Dialoggestaltung



**Abb. 12: Benutzer medizinischer Geräte**

Neben der Unterscheidung des Fachwissens und des Berufsstandes einer Gruppe ist es wichtig, auch innerhalb dieser Gruppen anhand weiterer Kriterien zu unterscheiden [284]. Ein Gerät, welches für die eine Nutzergruppe einfach und sicher zu verwenden ist, kann für Benutzer aus einer anderen Gruppe Schwierigkeiten verursachen. Beispielsweise könnte ein älterer Nutzer Schwierigkeiten haben, sich einzelne Sequenzen zu merken oder Aufgaben mit hohen Anforderungen an feinmotorische Fähigkeiten zu erfüllen, wogegen gut ausgebildetes und hoch motiviertes Personal mittleren Alters die gleichen Aufgaben ohne Probleme bewältigen kann [212].

Einige wichtige Charakteristiken einer Benutzerpopulation für den medizinischen Bereich gibt die amerikanische „Food and Drug Administration“ (FDA) vor [212]:

- Allgemeiner Gesundheits- und Geisteszustand (gestresst, entspannt, erholt, müde, beeinträchtigt durch Medikamente oder Krankheit) bei der Geräteverwendung,
- Physikalische Größe und Stärke,
- Sensorische Funktionen (Sehvermögen, Gehör, Haptik),
- Koordination (manuelles Geschick),
- Kognitive Fähigkeit und Speicher,
- Wissen über das verwendete Gerät und die damit verbundenen medizinischen Bedingungen,
- Geräteerfahrung (insbesondere ähnliche Geräte oder Benutzerschnittstellen),
- Erfahrungen mit der Funktionsweise des Gerätes,
- Motivation,
- Anpassungsfähigkeit an widrige Umstände.

Um sicherzustellen, dass die Anforderungen der Nutzer erfüllt werden, ist es notwendig, ihre Fähigkeiten, Ziele und Beeinträchtigungen zu erfassen, zu ver-

stehen und in die Produktentwicklung einzubringen. Generell ist festzuhalten, dass das medizinische Personal im Allgemeinen sehr leidensfähig ist. Kaum eine andere Berufsgruppe nimmt die in diesem Arbeitssystem vorherrschenden Belastungen wie Stress, Zeitdruck und Müdigkeit konsequenzlos hin [364]. Es kann aber auch davon ausgegangen werden, dass medizinisches Fachpersonal mit zu den am besten ausgebildeten und engagiertesten Personengruppen zählt. Daher sind auch meist nicht die Menschen das Problem bei Zwischenfällen, sondern das System, welches deutlich sicherer gemacht werden sollte [220]. Gerade dieser Berufsgruppe muss verständlich gemacht werden, dass in ihrem Arbeitssystem ein erhebliches Verbesserungspotential besteht, welches es gilt auszugleichen.

## 2 Ergonomie und Gebrauchstauglichkeit

In diesem Kapitel wird der für die Entwicklung der Standardprozedur elementare Bereich der Gebrauchstauglichkeit behandelt. Zuerst werden grundlegende Begriffe und Konzepte erläutert, um schließlich als Vorbereitung für die zu entwickelnde Standardprozedur die Gebrauchstauglichkeitsevaluationsmethoden vorzustellen und zu vergleichen. Am Ende dieses Kapitel werden die Gründe für eine Gebrauchstauglichkeitsbetrachtung erläutert, und es wird auf die Gebrauchstauglichkeitsbetrachtung in der Medizin eingegangen. Weiterhin werden die für diese Arbeit relevanten Gesetze, Normen und Richtlinien vorgestellt.

### 2.1 Grundlegende Begriffe und Konzepte

#### 2.1.1 Begriffsdefinition und Abgrenzung der Ergonomie und der Gebrauchstauglichkeit

Der Begriff Ergonomie stammt bereits aus dem 19. Jahrhundert (1857). „Ergonomie ist ein wissenschaftlicher Ansatz, damit wir aus dem Leben die besten Früchte bei der geringsten Anstrengung mit der höchsten Befriedigung für das eigene und das allgemeine Wohl ziehen“ [203]. Eine Renaissance erlebte die Ergonomie durch die Wiederentdeckung Murells im Jahre 1949 als Kennzeichnung einer neuen naturwissenschaftlichen Disziplin und durch die Gründung einer interdisziplinär tätigen Forschergruppe "Ergonomics Research Society"<sup>7</sup>.

Das Wort Ergonomie setzt sich aus den Worten ergon (griech.: Arbeit, Werk) und nomos (griech.: Gesetz, Regel) zusammen und beschreibt die Lehre von den Gesetzmäßigkeiten der menschlichen Arbeit. An zentraler Stelle dieser Wissenschaft stehen die Erforschung, die wissenschaftliche Analyse und die Verbesserung der Arbeitsbedingungen und der Mensch-Maschine-Schnittstelle (vgl. Kap. 1.2.1.2) [358]. Die Ergonomie stellt hierbei nicht die Arbeit, sondern den Menschen mit seinen Arbeitsaufgaben, seinem Umfeld und seinen Bedürfnissen in den Vordergrund.

---

<sup>7</sup> Gegründet 1949 als "Human Research Society", Namensänderung 1950 in "Ergonomics Research Society" und jetzt als „The Ergonomics Society“ bekannt.



Eine allgemeingültige Definition der Ergonomie liefert die Norm DIN EN 6385 [109]:

*„Wissenschaftliche Disziplin, die sich mit dem Verständnis der Wechselwirkungen zwischen menschlichen und anderen Elementen eines Systems befasst, und der Berufszweig, der Theorie, Prinzipien, Daten und Methoden auf die Gestaltung von Arbeitssystemen anwendet, mit dem Ziel, das Wohlbefinden des Menschen und die Leistung des Gesamtsystems zu optimieren.“*

Die Ergonomie ist demnach eine interdisziplinäre Wissenschaft, welche ihre Erkenntnisse aus den Bereichen der Ingenieurwissenschaften (Werkstoffkunde, Sicherheitstechnik), der Informatik (Interface Design), der Arbeitswissenschaften, der Arbeitsmedizin, der Anatomie (Anthropometrie, funktionelle Anatomie, Biomechanik), der Physiologie (Muskelfunktion, Kreisläufe, Sinne), der Neurophysiologie, der angewandten Psychologie, der angewandten Soziologie und der angewandten Wirtschaftswissenschaften gewinnt [343]. Angestrebte Ziele der Ergonomie sind nach Rohmert vornehmlich die Ableitung von Beurteilungskriterien hinsichtlich der Belastung und deren Auswirkungen, denen der Mensch ausgesetzt ist, und die Ableitung von korrektiven und konzeptiven, auf Normierung basierenden technischen und organisatorischen Arbeitssystemgestaltungen [343]. Besondere Berücksichtigung finden hierbei die Möglichkeiten, Erfahrungen und Einschränkungen des Menschen innerhalb der Arbeits- und Systemumgebung (vgl. Kap1.2.1.1) [33, 97, 105, 113, 426, 430].

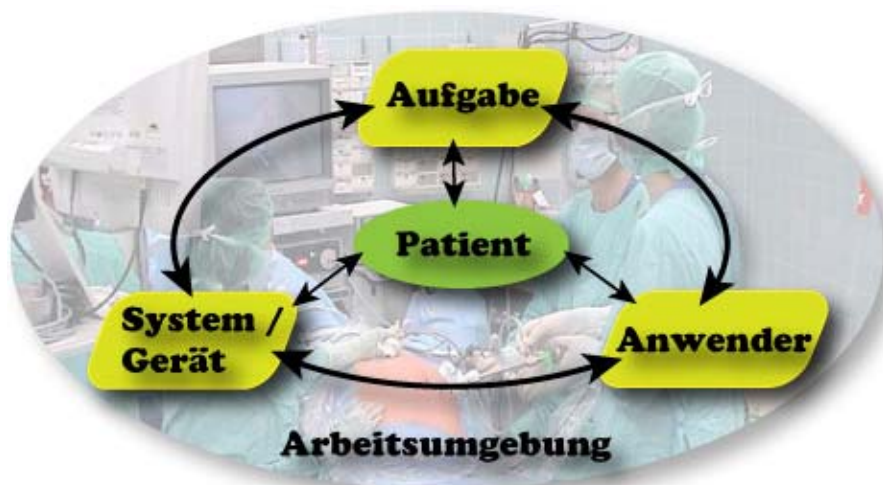
Im amerikanischen Sprachgebrauch wird Ergonomie meist als „Human Factors“ bezeichnet [78, 156, 264, 352]. Obwohl manchmal in der Literatur geringe Abweichungen zwischen Ergonomie und Human Factors zu finden sind, wie z.B. bei Hasler, der Human Factors als eine Erweiterung der Ergonomie um die Kognition bezeichnet [416], kann man dennoch von einer gleichen Bedeutung der beiden Begriffe ausgehen [78, 420].

Als zentraler Bestandteil der Ergonomie gilt die Gebrauchstauglichkeit (engl.: Usability). Oft wird auch unter den Begriffen Ergonomie und Gebrauchstauglichkeit dasselbe verstanden. Grundsätzlich richtig ist, dass Ergonomie und Gebrauchstauglichkeit das gleiche Ziel verfolgen, nämlich die Arbeitsumgebung und Produkte im Sinne des Anwenders möglichst verständlich und einfach zu gestalten. Im Gegensatz zur Ergonomie als übergeordnete Wissenschaft ist die Gebrauchstauglichkeit ein Attribut eines Gerätes oder Systems, welches eine Aussage darüber ermöglicht, inwiefern Nutzer mit einem bestimmten Gerät zu-

rechtkommen. Vom rein sprachlichen Gesichtspunkt aus bedeutet das: Ein Produkt „taugt zum Gebrauch“. Dies spiegelt sich auch in einer der ersten Definitionen für Gebrauchstauglichkeit, der seit 1966 existierenden Norm DIN 66050 als „die Eignung eines Gutes für seinen Verwendungszweck“ [92] wieder. Gebrauchstauglichkeit umfasst jedoch eine viel größere Bandbreite als in dieser Norm beschrieben. Heute weiß man, dass Gebrauchstauglichkeit in hohem Maße von dem bestimmten Nutzer, in einem bestimmten Nutzungskontext, mit seinen bestimmten Zielen abhängig ist und somit durch die Qualität des Zusammenspiels (Abb. 13) dieser Faktoren mit einem Produkt bestimmt ist [223]. Dementsprechend wird Gebrauchstauglichkeit in der Norm DIN EN 9241-11 als

*„das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufrieden stellend zu erreichen“*

definiert [113]. Um ein Ziel entsprechend dieser Definition erreichen zu können, muss ein Produkt nicht nur einfach zu verwenden sein, sondern auch eine ausreichende Funktionalität für die Aufgabenbewältigung bieten.



**Abb. 13: Abhängigkeiten im medizinischen Arbeitsumfeld**

Aus der genannten Definition ergibt sich, dass es nicht die absolute Gebrauchstauglichkeit für ein Produkt gibt. Die Gebrauchstauglichkeit ist immer abhängig von dem „bestimmten Benutzer“, der z.B. in Ausbildung, Kenntnisstand und Alter variieren kann, und von dem Nutzungskontext, in dem das Produkt verwendet wird.

Ein Vergleich der durch die DIN EN 9241-11 gelieferten Definition mit weiterer Literatur [22, 97, 105, 124, 223, 259, 284, 348, 349, 365, 371] zeigt, dass zwar unterschiedliche Begriffe zur Beschreibung der Gebrauchstauglichkeit verwendet werden, diese jedoch in ihrer Bedeutung ähnlich sind [223] und das gleiche Konstrukt mit unterschiedlichen Termini beschreiben (Tab. 3). Somit ist Gebrauchstauglichkeit die Eigenschaft eines Produktes, welche erst bei der Verwendung im Nutzungskontext durch einen bestimmten Benutzer bei einer bestimmten Aufgabe erkennbar wird. Sie stellt eine Nutzungsqualität und keine Produktqualität dar [223, 349].

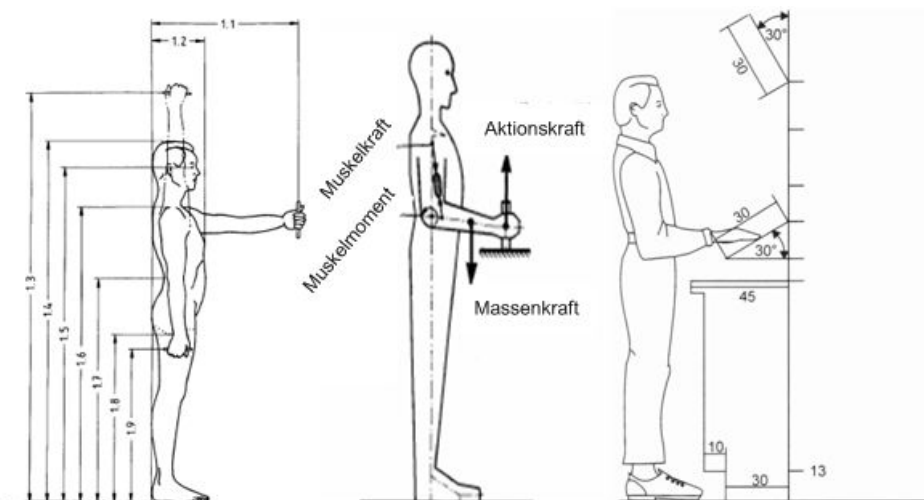
Vertreter	DIN EN 9241-11 [113]	Dumas/ Redish [124]	Mayhew [259]	Nielsen [288]	Rubin [348]	Shneiderman [368]
Benutzerleistung (objektiv)	Effektivität	Produktivität (Effektivität)	Einfache Anwendung	Fehlerrate	Einfache Anwendung	Fehlerrate
	Effizienz	Effiziente Aufgabenerledigung und einfache Anwendung	Einfache Anwendung	Effizienz	Einfache Anwendung	Schnelligkeit
			Einfach zu erlernen	Erlernbarkeit	Erlernbarkeit	Erlernbarkeit
				Einprägsamkeit		Einprägsamkeit
Benutzersicht (subjektiv)	Zufriedenstellung	Fokus auf den Benutzer (Zufriedenstellung)		Zufriedenstellung	Einstellung	Zufriedenheit

**Tab. 3: Exemplarischer Überblick über Gebrauchstauglichkeitsdefinitionen [223]**

Meistens wird der Begriff Gebrauchstauglichkeit bei interaktiven Systemen verwendet [113], er kann aber auch auf passivere Systeme übertragen werden [46, 50, 403]. Gebrauchstauglichkeit ist ein entscheidender Faktor bei der Gestaltung von Produkten, jedoch wird ihr derzeit noch nicht die Aufmerksamkeit zuteil, die üblicherweise anderen Faktoren, wie z.B. der Funktionalität gewidmet wird [6]. Damit ein Gerät als gebrauchstauglich anzusehen ist, muss es die Benutzer in dem jeweiligen Nutzungskontext bei der Aufgabenerfüllung derart unterstützen, dass sie ihre Ziele effektiv und effizient erreichen können [160]. Nutzer müssen bei ihrer Aufgabenerledigung zufrieden sein. Sie müssen sich darüber freuen, wie bequem und einfach sie ihre Ziele mit dem Gerät erreichen können [160]. Gebrauchstaugliche Geräte sind an den Nutzer angepasst und verlangen nicht von dem Nutzer, dass dieser sich an die Geräte anpasst.

### 2.1.2 Anthropometrie

Neben der Gestaltung der Benutzeroberfläche, der Gebrauchstauglichkeit und der genauen Betrachtung der Mensch-Maschine-Kommunikation fällt auch dem Bereich der Anthropometrie in der Ergonomie eine wichtige Rolle zu. Die Anthropometrie ist die Lehre von den Maßverhältnissen des menschlichen Körpers. Sie ist ein Wissenschaftsgebiet, das sich mit „dem Zusammenwirken von Mensch und Maschine befasst und die bestmögliche Gestaltung dieser Funktionseinheit hinsichtlich Leistung, Zuverlässigkeit, Wirtschaftlichkeit und Arbeitsbefriedigung durch Anpassung der Maschine an den Menschen zum Ziel hat“ [21]. Sie beschäftigt sich mit allen Messgrößen des Körpers oder seiner Teile, wie z.B. bei der statischen Anthropometrie mit Körpergröße, Körperbau und Körpergewicht, sowie bei der dynamischen Anthropometrie mit Körperkraft und Körperbewegung. Innerhalb der Ergonomie wird sie vor allem zur Gestaltung von Arbeitsplätzen, Werkzeugen und Inventar verwendet, sowie zur Untersuchung des Zusammenwirkens der Arbeitsmittel und der Messgrößen des Menschen (Abb. 14).



**Abb. 14: Beispiel einer anthropometrischen Betrachtung [228, 87, 89]**

Vielfältige ergonomische Beeinträchtigungen, die beispielsweise durch hohe sensomotorische und visuelle Anforderungen, Feinarbeit, statische Haltearbeit oder dynamische Muskelarbeit entstehen, können erhebliche Auswirkungen auf Gesundheit und Arbeitsleistung des Menschen haben. Neben temporären Druckstellen, Ermüdungen und dadurch bedingten Folgeerscheinungen können derartige Beeinträchtigungen beispielsweise zu langwierigen Muskel- und Skeletterkrankungen durch eine Überforderung des Bewegungsapparates führen [170]. Die Wichtigkeit der Anthropometrie wird verdeutlicht durch eine Betrachtung der

Statistiken: 30% der europäischen Arbeitnehmer, das sind ca. 40 Millionen Menschen, klagen über Rückenschmerzen, 17% berichten über Muskelschmerzen in Armen und Beinen und 45% geben an, in schmerzhafter oder anstrengender Stellung zu arbeiten [137]. Muskel- und Skeletterkrankungen gelten in den Mitgliedsstaaten der Europäischen Union als die häufigste Ursache für Arbeitsunfähigkeit. Laut dem Bundesverband der Betriebskrankenkassen (BKK) verursachten 2004 in Deutschland Muskel- und Skeletterkrankungen fast 27% aller krankheitsbedingten Ausfalltage der Beschäftigten [28].

Literatur über die Maße des menschlichen Körpers und anthropometrische Anforderungen ist vielfältig vorhanden. Grundlegende Maße sind bei Tillea & Dreyfuss [390] oder bei Lange & Windel [228] zu finden, sowie in den geläufigen Normen über die Maße des menschlichen Körpers [87, 88, 89, 110].

### **2.1.3 Gebrauchstauglichkeit – Aber wie?**

Eine wichtige Frage ist, wie die Gebrauchstauglichkeit eines Gerätes erreicht werden kann. Als oberste Prämisse hierbei gilt, entsprechend der DIN 9241-11, den Anwender, seine Arbeitsumgebung, Ziele und Aufgaben zu kennen und ihn in den Entwicklungszyklus mit einzubeziehen [33, 97, 105, 113, 426, 430], und zwar bevor erste Konzepte und Prototypen entstehen [19, 190]. Gebrauchstauglichkeit und damit verbundene Produktqualität beginnt bereits bei der Ideenfindung für ein Produkt und zieht sich durch den gesamten Entwicklungszyklus hindurch. Vorzüge, Verbesserungen, Versäumnisse und Mängel, welche in dieser Phase in die Entwicklung des Produktes mit hineinfließen, beeinflussen das Produkt ein Leben lang [361]. Deshalb ist ein erhöhtes Augenmerk auf die Gebrauchstauglichkeit des Produktes und auf die Anforderungen der Nutzer in dieser Phase besonders wichtig.

Als hilfreiche Unterstützung während der Geräteentwicklung sind Anforderungen an die Dialoggestaltung anzusehen, wie sie beispielsweise durch Shneiderman, Nielsen und die DIN EN 9241-110 beschrieben werden [114, 284, 371], wobei deren Einhaltung ein erhöhtes Maß an Gebrauchstauglichkeit garantiert. Nach dieser Norm handelt es sich hierbei um Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit und Lernförderlichkeit. Diese Kriterien gelten als ein Mindestmaß von Anforderungen und sollten für den Einzelfall weiter konkretisiert werden.

Zusätzlich stehen Entwicklern in der Regel eine Reihe von Styleguides<sup>8</sup> [145, 154, 374], Normen [99, 403] und Gestaltungsregeln [186, 349] zur Verfügung, welche entweder projektspezifisch entworfen [320] oder von anderer Stelle übernommen werden.

Unbedingt sollten bei der Gestaltung der Bedienoberfläche Gestaltungsgesetze, welche sich aus der Wahrnehmungslehre ableiten, beachtet werden. Diese Gesetze generieren sich aus der Wahrnehmung und Verarbeitung komplexer Sachverhalte oder Reize durch den Menschen. Die Umwandlung physikalischer Reize in Handlungsmuster erfolgt nach bestimmten Gesetzmäßigkeiten, die auf Erbstruktur beruhen, kulturell unabhängig sind und sich auf abstrakte Muster und Zusammenhänge beziehen [170, 222, 320, 349].

#### **2.1.4 Entwicklungsprozess der Gebrauchstauglichkeit**

*„Creating an interface is much like building a house:  
If you don't get the foundation right, no amount of  
decoration can fix the resulting structure” J. Raskin*

Wie bereits beschrieben, muss eine Gebrauchstauglichkeitsbetrachtung präventiv angewandt werden, indem sie bereits am Anfang der Produktentwicklung berücksichtigt wird, anstatt sie nachträglich für die Lösung von Problemen einzusetzen, wenn die Gestaltung von Geräten oder Arbeitssystemen bereits abgeschlossen ist. Die strukturierte Betrachtung von Gebrauchstauglichkeitsaspekten in der Produktentwicklung zum Zwecke der Entwicklung gebrauchstauglicher Benutzerschnittstellen nennt man im deutschen Sprachgebrauch „benutzerorientierte Entwicklung“, aber auch die internationalen Synonyme „Human Factors Engineering“ oder „User-Centered Design“ und deren Übersetzungen werden verwendet. Den größten Verbreitungsgrad hat der Begriff „Usability Engineering“ erlangt. Diese Disziplin ist nach Cakir die vorerst letzte Entwicklungsstufe im Bereich der Mensch-Maschine-Interaktion [50]. Sie ergänzt den „normalen“ Entwicklungsprozess um die Betrachtung des Faktors Mensch und durch die Bereitstellung von Methoden und Werkzeugen eines benutzerzentrierten Entwicklungsprozesses (Abb. 15), aus dem sich ergonomische Produktqualität ergibt. Als zentrales Vor-

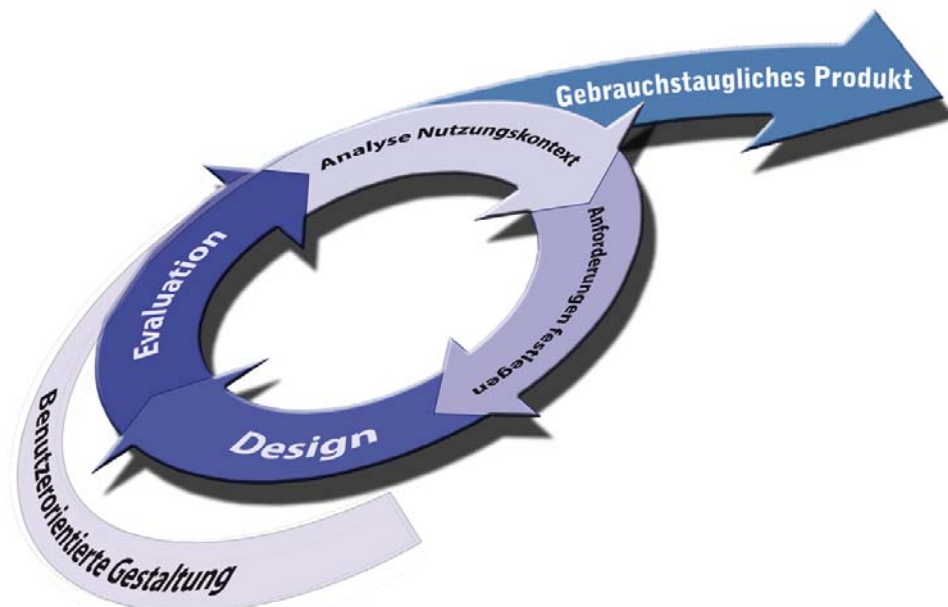
---

<sup>8</sup> Styleguides sind Regelwerke zur Gestaltung von Benutzeroberflächen. Sie beschreiben Anzeige- und Bedienelemente und geben Empfehlungen für deren Einsatz. Sie dienen somit der konsistenten Beschreibung von Bedienoberflächenmerkmalen innerhalb einer Anwendung oder über eine Anwendung hinaus, teilweise sogar auf globaler Ebene. Der Vorteil dieser Styleguides ist, dass ein Benutzer bei einer ihm bekannten Bedienoberfläche das Produkt leichter bedienen kann.

gehensmodell eignet sich der auf der DIN EN ISO 13407 basierende „benutzerorientierte Entwicklungsprozess“ [105]. Diese Norm gibt vier benutzerorientierte Gestaltungsaktivitäten vor, welche während einer Systementwicklung wiederkehrend auszuführen sind:

1. Verstehen und Festlegen des Nutzungskontextes
2. Festlegen von Benutzeranforderungen und organisatorischen Anforderungen
3. Entwerfen von Gestaltungslösungen
4. Beurteilen von Gestaltungslösungen gegenüber Anforderungen

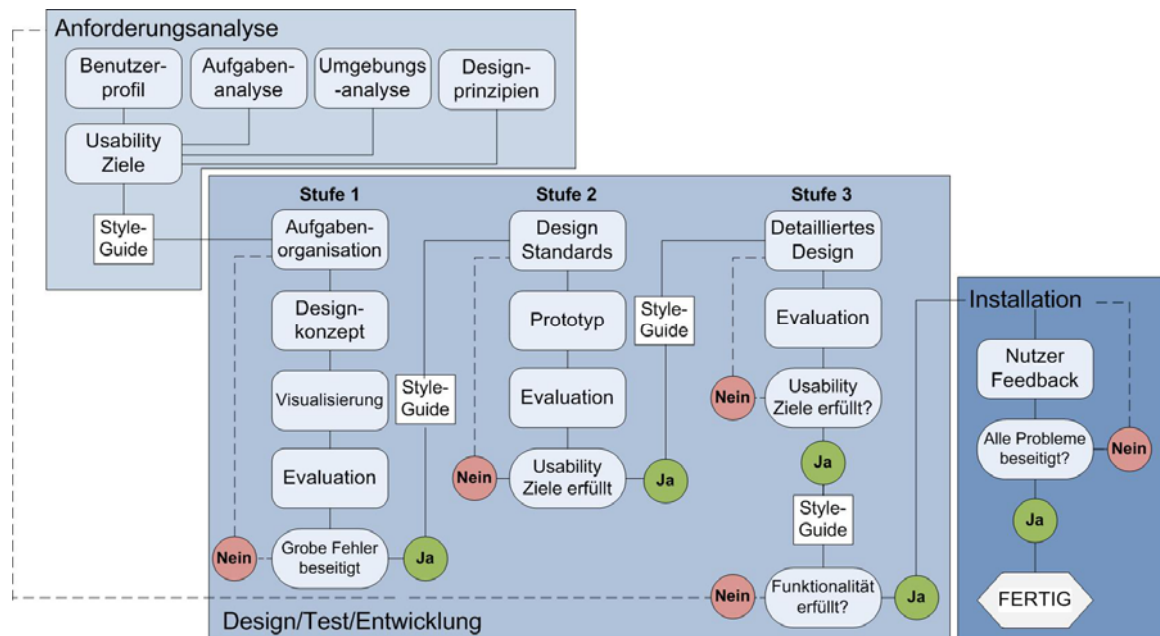
Dies dient dem Ziel, bereits während des Entwicklungsprozesses durch Orientierung an Kriterien der Gebrauchstauglichkeit und der Ergonomie sowie durch Beteiligung der Nutzer am Entwicklungsprozess ergonomische Mängel zu erkennen, zu beseitigen oder gar nicht erst aufkommen zu lassen.



**Abb. 15:** Ein benutzerzentrierter Designprozess wird solange durchgeführt, bis ein gebrauchstaugliches Produkt entwickelt ist [105].

Usability-Engineering setzt bei der iterativen Vorgehensweise des Software-Engineering an und ergänzt diese durch eine benutzerorientierte Herangehensweise, sowie durch Usability- und Ergonomiemodelle [105]. Neben dem in der DIN EN ISO 13407 beschriebenen nutzerzentrierten Entwicklungszyklus lassen sich noch weitere Entwicklungsprozesse finden. Sie alle besitzen neben der iterativen Vorgehensweise eine intensive Analysephase, in der die Anforderungen in Bezug auf den Benutzer und seinen Nutzungskontext gesammelt werden [114, 223, 259, 323, 346]. Ein weit verbreitetes Beispiel ist der „Usability Engineering

Lifecycle“ (Abb. 16). Mayhew teilt diesen Prozess in die Phasen „Anforderungsanalyse“, „Design, Test, Entwicklung“ und „Installation“ ein [259], unter Integration von bestehenden, bereits in der Praxis bewährten Designprinzipien und Styleguides [223, 259].



**Abb. 16: Usability Engineering Lifecycle nach Mayhew [259]**

Bei der gesamten Gebrauchstauglichkeitsbetrachtung ist zu berücksichtigen, dass von den derzeitigen ethischen, kulturellen bzw. religiösen Hintergründen und Erfahrungen der Nutzer ausgegangen wird. Eine für gut befundene Bedienungsschnittstelle kann in 10 Jahren u.U. schlechter bewertet werden, da sich die Erfahrungen der Nutzer und die Interaktionstechniken geändert haben. Beispielsweise könnte heutzutage eine Sprachbedienung für einige Nutzer noch irritierend wirken. In vielleicht fünf Jahren, wenn sich die Technik verbessert und die Sprachbedienung in den täglichen Gebrauch Einzug gehalten hat, könnte diese Art der Interaktion völlig akzeptiert werden.

Leider lässt die Praxis häufig ein Verständnis der Entwickler für den Nutzer und die Gebrauchstauglichkeitskultur vermissen. Wichtig ist, dass Gerätehersteller verstehen, dass sie ihre Nutzer kennen und sie als den Mittelpunkt des Entwicklungsprozesses ansehen müssen. Weiterhin ist ein iterativer Prozess durchzuführen, mit wiederkehrendem Design und Evaluationsphasen, denn keinem Entwickler gelingt die perfekte Gestaltung im ersten Versuch. Diesbezügliche Stolpersteine hat Galitz zusammengefasst [145]. Sie sind im Folgenden dargestellt:



- keine frühzeitige Analyse und Verständnis der Nutzerbedürfnisse und -erwartungen,
- Fokus auf die Verwendung von Designfeatures, die schön und prächtig sind („neat“ and „glitzy“),
- keine oder nur geringe Verwendung von Designprototypen,
- keine Gebrauchstauglichkeitsevaluation,
- schlechte Kommunikation zwischen den Mitgliedern des Entwicklungsteams,
- kein interdisziplinäres Team,
- keine iterative Entwicklung,
- Integration aller Systemkomponenten.

## 2.2 Ergonomie und Gebrauchstauglichkeit in der Medizin

Das eigentliche Ziel für den Einsatz von medizinischen Geräten ist es, die Gesundheit des Menschen wiederherzustellen bzw. zu erhalten. Neue Technologien unterstützen den Arzt in vielerlei Hinsicht und eröffnen ihm neue Behandlungsmöglichkeiten. Der technische Fortschritt führt dazu, dass immer mehr technische Geräte in die medizinische Arbeitsumgebung Einzug halten. Zwar werden hierdurch die Einsatz- und Anwendungsmöglichkeiten erhöht und die Qualität der Behandlung kann gesteigert werden, jedoch führt dieser stetige Zuwachs von Medizintechnik auch zu einer zunehmenden Anforderung an die Benutzer. Die medizinischen Arbeitsaufgaben werden mehr und mehr durch eine Interaktion mit der Technik geprägt, anstatt der Fokussierung auf den Patienten den Vorzug zu geben. Immer komplexere und kompliziertere Technologie beansprucht das Personal in erhöhtem Maße [42, 328]. Im medizinischen Bereich muss den Geräten viel Aufmerksamkeit geschenkt werden, die eigentlich dem Patienten gelten soll. Wenn der Benutzer überfordert ist, kann die Wahrscheinlichkeit von Fehlentscheidungen und Fehlbedienungen ansteigen [20]. Bereits heute hat ein Chirurg ähnlich viele Geräte zu bedienen wie ein Pilot im Cockpit [179, 233, 364]. Doch anders als am Arbeitsplatz Cockpit, an dem die Geräte aufeinander abgestimmte Systemeinheiten bilden und entsprechend den Bedürfnissen passend positioniert sind, finden sich im OP meist einzelne Elemente unterschiedlicher Hersteller mit jeweils eigenen Interaktionskonzepten wieder.

Es ist unbestreitbar, dass aus dem steigenden Technikeinsatz neue Anforderungen an Sicherheit, Kompatibilität, Systemunterstützung und vor allem an die Zusammenarbeit zwischen Mensch und Maschine und an die Ergonomie im medizinischen Bereich gestellt werden müssen. Gebrauchstauglichkeit gilt als

eine Voraussetzung für den sicheren und zuverlässigen Einsatz von Medizinprodukten [78]. Im Gegensatz zu anderen sicherheitskritischen Bereichen (wie z.B. Luft- und Raumfahrt, Atom- und Chemieindustrie, Militär) weist der medizinische Arbeitsplatz noch erhebliche Defizite bei der Beachtung von Gebrauchstauglichkeit und ergonomischem Design auf.

Die systematische und frühzeitige Betrachtung von Ergonomiegesichtspunkten, die Einbeziehung der Nutzer [366] in den Entwicklungszyklus und ein damit verbundenes ergonomisches, gebrauchstaugliches Design fördern eine sichere, zuverlässige und fehlerfreie Bedienung der Geräte. Besonders im Zusammenhang mit medizinischen Produkten ist es das Ziel, Gefahrenpotentiale zu verringern, die Verwendung medizinischer Geräte über den gesamten Produktlebenszyklus zu erleichtern und eine sichere und effektive Bedienung zu ermöglichen [212]. Bis vor Kurzem mussten jedoch medizinische Produkte lediglich Sicherheitsanforderungen hinsichtlich mechanischer und elektrischer Sicherheit erfüllen [362]. Die Berücksichtigung ergonomischer Gesichtspunkte muss erst seit 2006 durch die Einführung der DIN EN IEC 60601-1-6 (►Kap 2.3) nachgewiesen werden. Soweit es sich um ein medizinisch-elektrisches Gerät handelt, müssen Hersteller seit Einführung dieser Norm die Gebrauchstauglichkeit bei ihren Geräten nachweisen, um die für die Inverkehrgabe notwendige CE-Kennzeichnung zu erhalten.

Im Gegensatz zu den meisten Produkten werden an medizinische Geräte besondere Anforderungen gestellt, welche bezüglich Sicherheit vergleichbar sind mit den Anforderungen aus bekannten anderen sicherheitskritischen Bereichen. Hinzu kommt, dass medizinische Geräte in allen erdenklichen Situationen zum Einsatz kommen können. Ein Defibrillator wird beispielsweise in einem Krankenhaus zumeist bei guter Beleuchtung eingesetzt, muss aber auch dem Notarzt unter extremen Bedingungen in der Nacht zur Erstversorgung eines Unfallopfers dienlich sein. Eine weitere, medizinische Produkte betreffende Problematik ist die Kompatibilität der Geräte. Es kann nicht davon ausgegangen werden, dass bei der Benutzung immer gleiche Nutzungssituationen herrschen. Operationssäle werden für spezielle Operationen verschiedenartig aufgebaut. Ärzte fordern auf ihre Arbeitsweise abgestimmte Arbeitsumgebungen. Deshalb müssen die Geräte in unterschiedlichen Nutzungssituationen neben wechselnden Geräten einsetzbar sein. Im Gegensatz zu Flugzeugcockpits, Leitwarten etc. ist es selten möglich, ein komplett abgestimmtes System im OP zu installieren, da die situationsbedingte,

individuelle Gestaltung der meisten OPs gewährleistet bleiben soll. Trotzdem muss die Bedienung intuitiv und sicher sein. Dies gilt besonders für Geräte, welche selten verwendet werden. Hinzu kommt noch, dass medizinisches Personal in vielen Fällen unter enormem Zeit- und Leistungsdruck steht und eine hohe Verantwortung für Patient und Kollegen trägt. Die Arbeitsabläufe sind meistens sehr komplex und von Unterbrechungen durchzogen. Dieser enorme Stress kann schnell zu einer kognitiven Überlastung (engl.: cognitive load) und dann zu fehleranfälligen Situationen führen [191, 382], was durch eine komplizierte Bedienung medizinischer Geräte eher verstärkt bzw. gefördert, als, wie gewünscht, vermindert oder verhindert wird.

Aufgrund der sicherheitsbedingten Anforderungen an medizinische Geräte unterscheidet sich die Definition für Gebrauchstauglichkeit von anderen Industriezweigen [113] darin, dass neben den Maßen Effektivität, Effizienz und Zufriedenheit auch explizit die Sicherheit hervorgehoben wird. Die DIN EN IEC 60601-1-6 definiert Gebrauchstauglichkeit für die Medizin folgendermaßen:

*„Gebrauchstauglichkeit ist ein Maß für die Wirksamkeit, Effizienz und Zufriedenheit, mit der spezifische Anwender spezifische Ziele in einer bestimmten Umgebung innerhalb des Rahmens des bestimmungsgemäßen Gebrauchs /der Zweckbestimmung des Geräts erreichen. Viele dieser Faktoren können die Sicherheit in unterschiedlichem Umfang beeinflussen“ [97].*

Durch diese Formulierung wird ein konkreter Bezug zwischen Sicherheit und Gebrauchstauglichkeit hergestellt. Somit müssen Hersteller medizinischer Geräte einen Entwicklungsprozess für Gebrauchstauglichkeit durchführen, um die Sicherheit für Patienten, Anwender und andere Personen zu gewährleisten [97].

### **2.2.1 Notwendigkeit für Gebrauchstauglichkeit in der Medizin**

*“Medicine used to be simple, ineffective and relatively safe; now it is complex, effective and dangerous.“ C. Chantler*

Das medizinische Arbeitssystem ist mit einer ständigen Veränderung und Qualitätsverbesserung verbunden. Modernere und komplexere Geräte und Techniken halten fortwährend Einzug. Deutsche Medizintechnikunternehmen erzielen heute mehr als die Hälfte ihres Umsatzes mit Erzeugnissen, die weniger als zwei Jahre alt sind [20]. Anwender sind daher ständig mit neuen Geräten und unterschiedlichen Bedienkonzepten konfrontiert. Untersuchungen von Koneczny in drei Klini-

ken zeigen, dass in jedem OP-Saal durchschnittlich zehn Geräte stehen, die regelmäßig verwendet werden. Hinzu kommen 75 Geräte, die bei Bedarf hinzugeholt werden und sich aus 55 verschiedenen Gerätetypen von unterschiedlichen Herstellern zusammensetzen [221]. Längst vorbei sind die Zeiten, in denen eine ausgebildete Fachkraft nur ein System bedient. Vom OP-Personal wird erwartet, dass es mit allen Geräten sicher umgeht. Durch die Vielzahl unterschiedlicher Benutzerschnittstellen und Bedienkonzepte werden ausreichende Schulungsmaßnahmen unmöglich [408]. So ist es kaum verwunderlich, dass 97 Prozent der Chirurgen und OP-Pflegekräfte die Notwendigkeit sehen, den OP ergonomisch zu optimieren [256].

Matern et al. berichten, dass bei einer Befragung von 425 Chirurgen und 190 OP-Pflegekräften, 70 Prozent der Chirurgen und 50 Prozent der Pflegekräfte sich über Schwierigkeiten bei der korrekten Bedienung der Geräte äußerten. Dabei hatten über 40 Prozent der Befragten mehrfach Situationen erlebt, die für Personal oder Patienten gefährdend waren (Tab. 4) [256]. Bedienungsanleitungen für alle Geräte haben nach eigenen Angaben nur 6,7% der Chirurgen und 23,4% der OP-Pflegekräfte gelesen. Trotz der dem Anwender auferlegten Sorgfaltspflicht und der geforderten Einweisung [132, 275, 276] für jeden Anwender eines Medizinproduktes, sieht die Praxis meist anders aus. Es ist kaum möglich, dass das medizinische Personal alle Seiten der Gebrauchsanweisungen für sämtliche potentiell zu verwendenden Geräte liest und versteht. Angesichts der Vielfalt dieser Geräte mit ebenso umfangreichen Unterlagen ist der Lernaufwand erheblich. Oftmals besteht das Dokumentationsmaterial für nur ein Gerät aus mehreren DIN A4 Ordnern [201].

Aussage	Chirurgen (% von n = Antworten)	OP-Pflege (% von n = Antworten)
Geräte können nicht in jeder Situation intuitiv richtig bedient werden	69,8 % (n = 416)	48,9 % (n = 184)
Nicht ausreichend in der Gerätebedienung geschult fühlen sich	58,8 % (n = 410)	40,3 % (n = 186)
Die Gebrauchsanweisung für alle Geräte im OP haben gelesen	6,7 % (n = 418)	23,4 % (n = 188)
Probleme mit der Gerätebedienung haben mehrfach zur Gefährdung der im OP befindlichen Personen geführt	39,7 % (n = 388)	47,7 % (n = 176)

Tab. 4: Aussagen zur Gerätebedienung von Chirurgen und OP-Pflegekräften [256]

Durch die einerseits hohe Fluktuation von Bedienern in den einzelnen Abteilungen und die andererseits geforderte Flexibilität der Anwender ist es praktisch kaum vorstellbar, dass alle Anwender eine Einweisung für jedes potentiell zu verwendende Gerät bekommen haben. Zahlen von Koneczny belegen, dass ca. ein Drittel (34,5%) aller Befragten keine Einweisung in die zu verwendenden Geräte erhalten haben [221]. Aber auch eine ausreichende Einweisung ist kein Garant für Fehlerfreiheit. Eine im Rahmen dieser Arbeit durchgeführte Studie zeigt, dass es bei Pflegepersonal und Auszubildenden zum Anästhesietechnischen Assistenten in einem Nutzertest eine Woche nach einer Einweisung immer noch bei den meisten Primäraufgaben zu Fehlern kam, welche teilweise sogar sicherheitskritisch waren (► Kap. 4.2.3). Auch aus diesem Grund ist Weinger zuzupflichten, dass Hersteller möglichst einfache und intuitiv zu bedienende Produkte entwickeln sollten, die auch ohne Training bedient werden können [408].

Mit steigender Anzahl immer komplexer werdender Geräte und erhöhter physischer Belastung nimmt auch das Risiko von Benutzerfehlern zu [6, 97, 414, 434]. Ein Großteil dieser Fehler ist auf mangelnde Gebrauchstauglichkeit zurückzuführen. In komplexen Arbeitssystemen ist davon auszugehen, dass 70 bis 80 Prozent aller Fehler auf menschliches Versagen zurückgeführt werden können [80, 186, 220]. Untersuchungen zeigen [139] in Übereinstimmung mit der Expertenmeinung, dass hiervon bis zu 60 Prozent Probleme der Mensch-Maschine-Kommunikation sind [11, 30, 78, 80, 187, 256]. Auch eine Ansammlung von relativ geringfügigen Beeinträchtigungen der Benutzer-Gerät-Schnittstelle kann die Wahrscheinlichkeit von Benutzerfehlern mit resultierenden Konsequenzen erhöhen [97]. Bleyer und auch von der Mosel geben an, dass etwa 60 Prozent der Vorkommnisse, die mit medizintechnischen Geräten in Zusammenhang stehen, auf Missverständnissen zwischen Mensch und Maschine beruhen [30, 405]. Beeinträchtigungen führen zu zunehmender Arbeitslast. Unter damit verbundene Stress nimmt die Wahrscheinlichkeit von Fehlern unterschiedlichen Typs zu (Tab. 5). Es steht weniger Zeit zur Verfügung, um die Arbeitsschritte zu überprüfen bzw. um Maßnahmen zu überlegen, und es erwächst dadurch eine große Wahrscheinlichkeit für Ungenauigkeiten und grobe Fehler [97].

Stress	Erhöhung der Fehlerwahrscheinlichkeit	
	Erfahrene Personen	Unerfahrene Personen
Sehr niedrig	mal 2	mal 2
Ziemlich hoch	mal 2	mal 4
Extrem hoch	mal 5	mal 10

Tab. 5: Fehlerwahrscheinlichkeit in Abhängigkeit von Stress und Arbeitserfahrung [20]

Eine Untersuchung von Nielsen zeigte, dass viele Probleme im Krankenhaus als klassische Gebrauchstauglichkeitsprobleme schon lange bekannt sind [286]. Hierzu gehören: Irreführende Standardwerte, das Versäumnis neue Befehle gegen vorhergehende zu prüfen, schlechte Lesbarkeit, geistige Überbelastung, Datumsfehler und übermäßig komplizierter Workflow. Weitere Bedienfehler, die zu Unfällen führen, sind laut VDE (Verband der Elektrotechnik, Elektronik und Informationstechnik e.V.) beispielsweise: Verwechslung von Anschlüssen, Lesefehler von Displayangaben aufgrund mangelhafter Darstellung, Fehlbedienung aufgrund nicht vorhandener Funktions- und Positionsanzeigen und Mangel an intuitiver Bedienführung vor allem bei selten benutzten Geräten [363], sowie nach der DIN EN ISO 14971 fehlerhafte Darstellung, unklare Zustände, ungenügende Sichtbarkeit, Hörbarkeit und Berührbarkeit [108].

Leider rangiert das Gesundheitswesen in Bezug auf Risikobewusstsein immer noch weit hinter der Industrie, in welcher Risikobetrachtung und -vermeidung sowie konsequente Investition in sichere Systeme nicht mehr wegzudenken sind [220]. Zwar hat sich in den letzten Jahren durch gesetzliche Bestimmungen und Normen einiges in diesem Bereich geändert, doch haben viele Hersteller diese noch nicht vollständig umgesetzt. Dennoch ist ein Anfang gemacht, und es gilt, diesen weiterzuentwickeln und ein Bewusstsein für Gebrauchstauglichkeit sowohl bei Herstellern und Betreibern als auch bei Anwendern zu entwickeln [156]. Bei Letzteren ist vermehrt zu beobachten, dass sie sich mit schlechtem Design zufriedengeben und selbstständig Abhilfen (sog. „Work Arounds“) (Abb. 17) entwickeln, mit denen sie glauben, die Probleme umgehen zu können [156]. Ihnen muss klar gemacht werden, dass Alternativen möglich sind.



**Abb. 17: Ein Hochfrequenz-Chirurgiegerät benötigt zusätzliche Informationen**

Die richtige Auswahl von Systemen ist mittlerweile für jedes Unternehmen erfolgskritisch. Qualitätsvergleiche und Kaufentscheidungen finden aber im Ge-

sundheitssystem vorzugsweise nur nach ökonomischen und funktionalen Produktmerkmalen statt. Darüber wird die Bedienbarkeit und Benutzerfreundlichkeit der Geräte meist vergessen. Krankenhausbetreiber müssen daher erkennen, dass die Einbeziehung von Gebrauchstauglichkeitsgesichtspunkten in die Kaufentscheidung notwendig ist.

Gerätehersteller werden durch den Druck des Marktes dazu getrieben, ihre Geräte mit Extrafunktionen (Features) auszustatten und vergessen dabei die Faustregel, dass für 20% der Gerätefunktionen 80% der Zeit verwendet wird. Laut einer Studie von Wiklund forderten lediglich 11% von 36 Intensivkrankenschwestern mehr Features bei den Geräten. 33% meinten bereits im Jahr 1992, dass die Geräte zu viele Extrafeatures hätten. 44% von ihnen fühlten sich nicht ausreichend in der Gerätebedienung geschult. Im Allgemeinen waren die Krankenschwestern besorgt um die Gebrauchstauglichkeit medizinischer Geräte und forderten von den Herstellern, sich mehr mit diesem Thema zu befassen. Besonders vor dem Hintergrund, dass ihre Arbeitslast zu wenig Zeit für die Erlernung eines Gerätes erlaubt, sollten vor allem Primäraufgaben medizinischer Geräte (vgl. Kap. 1.2.2.2) intuitiv und leicht zu erlernen sein [414].

Die medizintechnischen Unternehmen müssen aber auch für Veränderungen bereit sein. Um mögliche Probleme abzuwenden, bestehende Probleme zu beheben, Fehler zu vermeiden, ein generell höheres Maß an Kundenzufriedenheit zu erreichen und die Patientensicherheit zu erhöhen, ist für medizinische Unternehmen die Verwendung eines benutzerzentrierten Designprozesses (vgl. Kap. 2.1.4) zwingend notwendig. Dieser Prozess unterstützt die Beteiligten des Gesundheitssystems dabei, Probleme zu analysieren und praktikable und wirksame Gegenmaßnahmen zu ergreifen [156]. Hierdurch entstehen Geräte, welche die Nutzer gerne und sicher in realistischen Bedingungen verwenden können.

Aber auch in den „best designed“ Systemen mit höchst motivierten und trainierten Anwendern werden weiterhin Fehler gemacht. Deshalb fordert Norman von den Herstellern möglichst fehlerunempfindliche Systeme und eine Reduzierung der Fehlerauswirkung [294]. Systeme müssen so entwickelt werden, dass sie präventiv Fehler vermeiden und im Falle dennoch auftretender Fehler die Auswirkungen verringern. Eine ähnliche Anforderung an die Wechselwirkung zwischen Bedienperson und Maschine ist in der DIN EN 614-1 festgelegt: „Maschinen müssen fehlertolerant sein und der Bedienperson Maßnahmen für den Umgang mit Fehlern vorgeben“ [99].

## 2.2.2 Gründe für eine Gebrauchstauglichkeitsbetrachtung

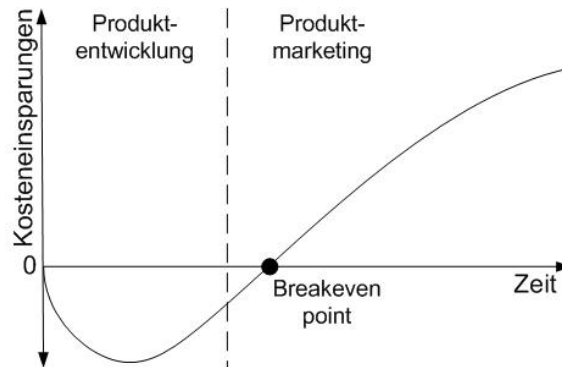
Wer kennt nicht die Frustration bei der Verwendung von technischen Geräten, wenn diese nicht so wollen, wie man selbst. Bei Geräten, welche durch einen benutzerorientierten Entwicklungsprozess (vgl. Kap. 2.1.4) gebrauchstauglich gestaltet sind, ist dieser Frustfaktor weitestgehend reduziert. Derartige Produkte zeichnen sich durch einfache Verwendung und schnelle Erlernbarkeit aus und lassen den Nutzer so die angestrebten Ziele mit einer hohen Zufriedenheit effektiv und effizient erreichen. Fehler treten seltener auf, und im besten Fall macht sogar die Arbeit mit den Geräten Spaß, was die Motivation deutlich erhöht. Die Beachtung von ergonomischen Aspekten und der Gebrauchstauglichkeit legitimiert sich jedoch nicht allein durch die gesteigerte Anwenderzufriedenheit und den effektiven und effizienten Umgang mit den Geräten, auch ökonomische Einsparpotentiale auf Seiten der Hersteller, Anwender und Betreiber, sowie gesteigerte Patienten- und Anwendersicherheit spielen eine Rolle.

### 2.2.2.1 Ökonomischer Hintergrund

Kostenersparnis wird gerade im Gesundheitssystem immer wichtiger. Nicht erst durch die Einführung der DRGs (vgl. Kap. 1.2.2.1) ist der medizinische Arbeitsablauf einem erhöhten Leistungs- und Effizienzdruck ausgesetzt. Die Behandlung von Patienten ist einerseits mit einem hohen Kosteneinsatz verbunden, andererseits müssen Gesundheitsfürsorger immer wirtschaftlicher handeln. Der Nutzen gebrauchstauglicher Qualität, verbunden mit Kosteneinsparungen, tritt auf sehr vielfältige Weise zutage. Positiveffekte durch gebrauchstaugliche Geräte sind sowohl auf Seiten der Nutzer als auch auf Seiten der Hersteller beobachtbar.

Der vielleicht für den **Produkthersteller** wichtigste Faktor für Usability Engineering ist die Wirtschaftlichkeit von Gebrauchstauglichkeitstests. Für Unternehmen sind diese Tests oft Ereignisse, die mit viel Aufwand und hohen Kosten verbunden sind [240]. Dagegen ist Usability Engineering eher als eine Investition anzusehen, da für Unternehmen die monetären Vorteile, welche durch Einbeziehung ergonomischer Betrachtung und durch einen nutzerzentrierten Designprozess entstehen, gegenüber den anfallenden Kosten in der Regel überwiegen [211, 260]. Wiklund veranschaulicht in Abb. 18 den Verlauf der zu erwartenden Kosteneinsparung durch Usability Engineering für ein hypothetisches Produkt. Erkenntlich wird, dass relativ schnell nach der Produkteinführung die Kosteneinsparungen größer als die getätigten Investitionen sind [414].





**Abb. 18: Kosteneinsparungen durch Usability Engineering**

Für Hersteller ist der größte Benefit der Gebrauchstauglichkeitsbetrachtung ein intuitiv einfach zu verwendendes Produkt, welches an die Zweckbestimmung angepasst ist, den Kundenbedürfnissen entspricht und den Benutzer zufrieden stellt. Gebrauchstaugliche Produkte, welche auf den Benutzer zugeschnitten sind, haben auf dem Markt Erfolg, weil sie die Erwartungen der Kunden erfüllen, einfacher zu bedienen sind, öfter verwendet werden und eine größere Zielgruppe ansprechen [39]. Die Gestaltung der Benutzeroberfläche ist entscheidend für die Akzeptanz und Qualität [22], sie führt zu Kundenbindung und Vertrauen. Zusätzlich können auch Kosten in der Entwicklung eingespart werden. Dadurch, dass Entwickler mit Verfolgung des Usability Engineerings nach Analyse der Kundenbedürfnisse, der Nutzungssituation, der Nutzeraufgaben und der Nutzungsumgebung die Anforderungen kennen, welche an das Produkt gestellt werden, wissen sie frühzeitig, wie das Produkt aussehen soll. Auch dient ein benutzerfreundliches Produkt als Differenzierungs- und Positionierungsinstrument auf dem Markt. Das Usability Engineering liefert eine Fülle an Methoden, welche Systeme mit hohem ergonomischem Wert entsprechend der Ziele und Arbeitsaufgaben der Nutzer entstehen lassen.

Gerade in sicherheitskritischen Systemen ist das mit unergonomischen und nicht gebrauchstauglichen Geräten verbundene Risiko von elementarer Bedeutung. In Fällen, in denen es um die Sicherheit von Menschen geht, kann bei Geräteproblematiken die Existenz des Herstellers auf dem Spiel stehen. Daher darf er sich bei sicherheitskritischen Systemen keinen Fehler erlauben, der durch schlecht zu bedienende Geräte entstünde. Aber auch ohne diese extreme Konsequenz leidet der Ruf des Herstellers. Hinzu kommt noch, dass Ressourcen bei der Fehlerbeseitigung gebunden werden. Statt sich neuen Projekten zu widmen, müssen erst die Forderungen von Geschädigten erfüllt werden [361]. Viel besser als nachträg-

liche Korrekturmaßnahmen ist eine Einbeziehung der Nutzer bereits zu Beginn des Entwicklungsprozesses, um eventuelle Fehlentwicklungen frühzeitig erkennen und beheben zu können.

Nach einer Studie der Standish Group werden nur 9-16% der Softwareentwicklungsprojekte pünktlich und im Rahmen des Budgets fertiggestellt, mehr als 30% der Projekte werden vor Abschluss abgebrochen. 52% der Projekte benötigen die zwei- bis dreifache Entwicklungszeit der zu Beginn veranschlagten. Ursache hierfür sind in erster Linie fehlende Informationen über die Benutzer und die Anforderungen, die an das Produkt gestellt werden [400]. Bei Befragungen der Softwaremanager über den Grund sind die Antworten eng mit mangelndem oder unzureichendem Usability Engineering verbunden. Hierzu gehören häufige Änderungswünsche der Nutzer, nicht erkannte Arbeitsaufgaben, geringes Verständnis der Nutzeranforderungen und unzureichende Benutzeranalysen [284]. Gerade diese Probleme können durch eine konsequente Einbeziehung der Nutzer in den Entwicklungsprozess mit Hilfe von Gebrauchstauglichkeitsmethoden vermieden oder zumindest verringert werden. In diesem Bereich werden Aufwand-Nutzen-Relationen von bis zu 1:100 erzielt [6, 211, 316]. Wixon & Jones dokumentierten eine 80-prozentige Umsatzsteigerung eines Produktes, welches zuerst ohne die Beachtung von Gebrauchstauglichkeit hergestellt wurde. Käufer gaben explizit die Gebrauchstauglichkeit des Produktes als Kriterium für die Kaufentscheidung an [211].

Für den Hersteller fällt ein hoher Kostenanteil erst nach dem Projektabschluss an. Pressmann gibt an, dass 80% der Kosten eines Softwareentwicklungszyklus in der Wartungsphase anfallen [316]. Die Behebung eines Problems nach einer Produktfreigabe ist meist schwierig und kostet um ein Vielfaches mehr als in früheren Stadien [258]. Dies liegt daran, dass zu Beginn des Designprozesses zentrale Designentscheidungen getroffen werden, welche die Produktkosten und die Leistung des Produktes bestimmen (Abb. 19). Nach dem „U.S. Department of Health & Human Services“ machen Fehler und Unzuverlässigkeit der Produkte nur 20% der anfallenden Wartungskosten aus, der größte Teil (80%) ist auf nicht erfüllte oder nicht vorhergesehene Nutzeranforderungen zurückzuführen [395].

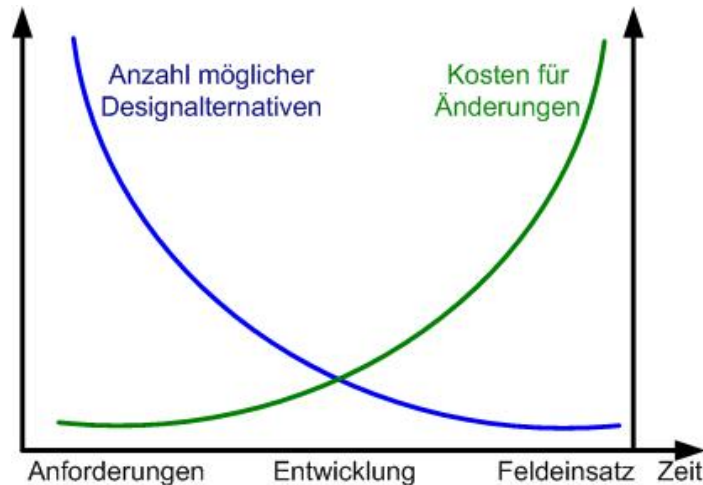


Abb. 19: Kosten im Entwicklungsprozess [in Anlehnung an: 134]

Das global agierende IT- und Beratungsunternehmen IBM schätzt, dass für jeden Dollar, welcher in die Benutzerfreundlichkeit von Anwendungen investiert wird, das Zehnfache zurückfließt [248]. Die Firma Sun Microsystems investierte ca. \$20.000 und konnte hierdurch Einsparungen von \$152 Millionen erreichen [252]. Boeing konnte durch die Beachtung ergonomischer Gesichtspunkte und benutzergerechter Cockpitausstattung die Pilotenanzahl in der Boeing 757 von drei auf zwei reduzieren [284]. Diese Aspekte zeigen, dass Hersteller die Kosten für Usability Engineering nicht scheuen, sondern vielmehr als langfristige und nachhaltige Investitionen in die Entwicklung benutzerfreundlicher Produkte sehen sollten.

Auf **Nutzerseite** (direkte Anwender oder Unternehmen, in denen das Produkt verwendet wird) sind durch gebrauchstaugliche Produkte zum einen Kostenersparnisse dank einer erhöhten Produktivität durch bessere Aufgabenerfüllung und zum anderen Kostenersparnisse durch geringeren Trainingsaufwand und weniger Fehler möglich [39]. Durch eine Verminderung der durch Gebrauchstauglichkeitsprobleme entstehenden Fehler kann die Performance enorm gesteigert werden. Shneiderman gibt an, dass bereits uneinheitliche Farbgebung, Größe oder Anordnung von Drucktasten das Arbeitstempo um 5-10% verlangsamen und verwirrende Begriffe auf Eingabeteilen zu Verzögerungen von bis zu 25% führen können [251, 369]. Arnold & Roe ermittelten, dass 30 bis 40% des gesamten Interaktionszeitraums für Fehlerbehebung anfallen. Zusätzlich haben Fehler in den meisten Fällen negative emotionale und demotivierende Folgen [8]. Mangelnde Gebrauchstauglichkeit bringt neben erhöhtem Risiko und damit verbundener Patientengefährdung verlängerte Arbeitszeiten, größere Personalkapazitäten, höheren Schulungsaufwand, stärkere Frustration und eine geringere Gerätever-

wendung durch nachlassendes Vertrauen mit sich. Weniger Fehler und effektivere Bedienung der Geräte helfen somit auf Seiten der Anwender Zeit und damit auch Kosten einzusparen. Wenn beispielsweise eine früher 5 Minuten dauernde Aufgabe bei einem mit Usability Engineering entwickelten Produkt nur noch 4 Minuten dauert, und diese Aufgabe in einem Krankenhaus von 30 Personen durchschnittlich 5 mal am Tag durchgeführt wird, können bei 200 Arbeitstagen ( $1 \times 30 \times 5 \times 200 = 30.000$  Minuten) 500 Arbeitsstunden eingespart werden.

Für den medizinischen Bereich fordert das englische „Department of Health“ ausdrücklich mehr anwenderorientierte Geräte, um Benutzerfehler zu vermeiden. In England passieren jährlich ca. 850.000 „Zwischenfälle“ aufgrund von Benutzerfehlern. Die Kosten hierfür werden auf £400 Millionen geschätzt [186]. Für die USA liefert das „Institute of Medicine“ ähnliche Zahlen. Es geht davon aus, dass die nationalen Kosten, welche aufgrund vermeidbarer medizinischer Fehler anfallen, zwischen 17 und 29 Milliarden Dollar liegen [220]. Bringt man diese Schätzungen mit der Annahme von Bleyer zusammen, dass ca. 60% der Vorkommnisse, die mit medizintechnischen Geräten in Zusammenhang stehen, auf Missverständnissen zwischen Mensch und Maschine beruhen [30], so liegen die in den USA entstehenden Kosten für vermeidbare medizinische Fehler in der Mensch-Maschine-Kommunikation zwischen 10,2 und 17,4 Milliarden Dollar. Für Deutschland berechnet Backhaus die jährlichen Kosten im Bereich der Intensivmedizin aufgrund schlechter Gebrauchstauglichkeit der Geräte auf ca. 396 Millionen Euro, was 3,7% der jährlichen Gesamtaufwendungen für diesen Sektor ausmacht [11]. Von gebrauchstauglichen Geräten profitieren sowohl Anwender als auch Betreiber. Problematisch ist jedoch für die Anwender, diese Geräte zu erkennen. Gängige Praxis bei der Produktanschaffung im Krankenhaus ist eine Probestellung (► Kap. 3.2). Hierbei wird jedoch in der Regel nicht speziell und intensiv genug auf Gebrauchstauglichkeit und Ergonomie geachtet. Meistens wird von der Funktionalität des Gerätes auf die Leistungsfähigkeit geschlossen. Viele Anwender, die leicht und intuitiv zu bedienende Geräte erwarten, erkennen jedoch die Bedienproblematik häufig erst im praktischen Einsatz [280]. Irreführend ist vielfach, dass das Prädikat „ergonomisch“ oder „gebrauchstauglich“ oft zu verkaufsförderndem Marketing missbraucht wird, ohne dass Studien die ergonomischen Eigenschaften der Produkte belegen [255]. Bisher gibt es noch kein speziell geschütztes Siegel, welches nach aussagekräftigen Gebrauchstauglichkeitsuntersuchungen dem Kunden ein gebrauchstaugliches Produkt ausweist.

Beim Hersteller	Beim Anwender
<ul style="list-style-type: none"> <li>➤ Gesteigerte Kundenzufriedenheit</li> <li>➤ Vertrauen und Loyalität der Kunden</li> <li>➤ Kundenbindung</li> <li>➤ Geringere Wartungs- und Supportkosten</li> <li>➤ Geringerer Schulungsaufwand</li> <li>➤ Höhere Gewinne</li> <li>➤ Geringere Entwicklungskosten</li> <li>➤ Höhere Marktakzeptanz</li> <li>➤ Bessere Wettbewerbsfähigkeit</li> <li>➤ Entwicklung eines an den Nutzer und seine Aufgaben angepassten Produktes</li> <li>➤ Kürzere Entwicklungszeiten</li> <li>➤ Schnellere Marktreife</li> <li>➤ Einsparungen durch frühzeitige Änderungen im Produktentwicklungszyklus</li> </ul>	<ul style="list-style-type: none"> <li>➤ Zufriedenheit</li> <li>➤ Gesteigerte Effektivität und Effizienz</li> <li>➤ Geringerer Schulungsaufwand</li> <li>➤ Schnellere Einarbeitungszeit</li> <li>➤ Erhöhte Produktivität</li> <li>➤ Höhere Sicherheit</li> <li>➤ Weniger Fehler</li> <li>➤ Geringere Geräteeinsatzzeiten</li> <li>➤ Schnellere Aufgabenerfüllung</li> <li>➤ Kostenersparnisse</li> </ul>

**Tab. 6: Nutzen durch gebrauchstaugliche Geräte für Anwender und Hersteller**

Leider ist vielen Anwendern und Unternehmen der Benefit, welcher durch Usability Engineering entsteht (Tab. 6), noch nicht bekannt. Solange Anwender diesen Vorzug nicht erkennen, und Hersteller medizinischer Geräte Gebrauchstauglichkeit als eine vom Gesetzgeber auferlegte lästige Pflicht auffassen, besteht nach Ansorge et al. die Gefahr der Marginalisierung von entsprechenden Gebrauchstauglichkeitsprüf- und -verbesserungsverfahren [6].

### 2.2.2.2 Anwender und Patientensicherheit

Das wohl im medizinischen Bereich wichtigste Argument für gebrauchstaugliche Geräte ist die Anwender- und Patientensicherheit. Durch schlecht gestaltete Benutzerschnittstellen und Probleme in der Mensch-Maschine-Kommunikation sind Fehler möglich, welche Auswirkungen auf die Gesundheit des Anwenders oder des Patienten haben [346, 366]. Selbst erfahrene Nutzer werden von irreführenden oder unlogischen Benutzerschnittstellen beeinflusst [352]. Die Bedienprobleme der Menschen, welche zu Fehlern führen, werden oft durch die Technik unterstützt oder sogar provoziert [220]. Hierbei sind Auswirkungen von geringen Folgen, wie beispielsweise eine Ermüdung der Augen des Anwenders, bedingt durch einen schlechten Kontrast des Anzeigesystems, bis hin zum Tod des Patienten möglich [80, 97, 352, 366].

US-amerikanische Studien belegen, dass medizinische Fehler zu den acht häufigsten Todesursachen in den USA zählen [220]. In dem im Jahre 1999 veröffentlichten Report des „Institute of Medicine“ (IOM) schätzen Kohn et al., dass mehr

als die Hälfte der medizinischen unerwünschten Ereignisse aufgrund vermeidbarer medizinischer Fehler passieren. Unter der Annahme von jährlich 33,6 Millionen Einweisungen in US-amerikanische Krankenhäuser und unter Berufung auf Studienergebnisse in Colorado und New York, dass jährlich 44.000 bis 98.000 Amerikaner durch medizinische Fehler sterben, wird der daraus resultierende Anteil innerhalb der Todesursachen deutlich [220]. Weitere Studien in Australien, Kanada und England zeigen ähnliche Ergebnisse [14, 404, 418]. Das Aktionsbündnis für Patientensicherheit zieht nach Untersuchungen von 51 nicht länderbezogen verteilten Studien die Schlussfolgerung, dass diese Ergebnisse auf Deutschland übertragbar sind, und somit die Sterblichkeit von Krankenhauspatienten aufgrund eines vermeidbaren unerwünschten Ereignisses mit einer Größe von 0,1 Prozent nicht überschätzt wird [359].

Zwar haben seit Veröffentlichung des IOM Reports die Themen „Fehler in der Medizin“, „Patientensicherheit“ und „Gebrauchstauglichkeit im medizinischen Kontext“ in der Literatur zunehmendes Interesse erlangt, und es sind bereits kleine, aber wichtige Änderungen zur Verbesserung der Patientensicherheit in Krankenhäusern umgesetzt worden, jedoch schreitet dieser Umdenkungsprozess zu langsam voran, und es ist noch lange kein zufriedenstellendes Level erreicht [230]. Immer noch wird im Gesundheitswesen eine zu hohe Anzahl an Patienten unnötig verletzt oder stirbt durch vermeidbare Fehler. Beispielsweise werden ca. 10% der in englische Krankenhäuser eingewiesenen Patienten durch Fehler während der Behandlung verletzt. Dies entspricht etwa 850.000 Schadensfällen pro Jahr [406].

Eine Studie von Montag et al. analysiert 1.330 Vorkommnismeldungen für den Bereich OP zwischen 2000 und 2006 beim Deutschen Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). 48% der Vorkommnisse sind hierbei auf Anwender-<sup>9</sup> und Anwendungsfehler<sup>10</sup> zurückzuführen, wovon 19% reine Anwendungsfehler sind. 37% der Anwendungsfehler treten beim Einsatz im OP auf. 97% der Anwendungs- und Anwenderfehler sind mit einer Gefährdung der im OP befindlichen Personen verbunden (90% Gefährdung der Patienten, 7% Gefährdung der Mitarbeiter) [273]. Erschreckend hierbei ist, dass dies nur in einem

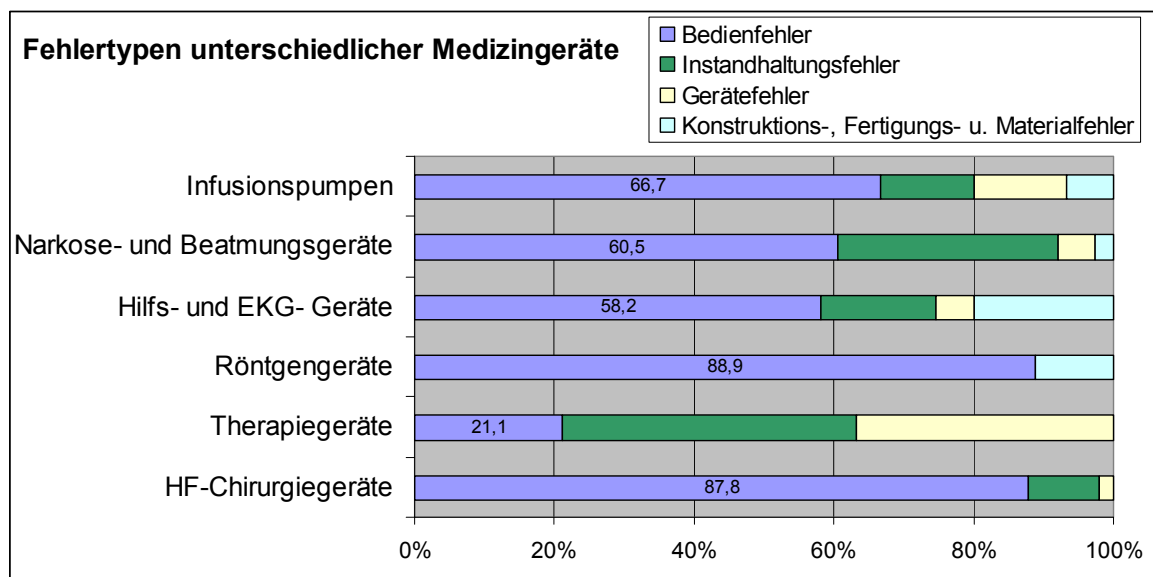
---

<sup>9</sup> Anwenderfehler: Fehler, bedingt durch fehlende fachliche Kenntnisse, Verletzung der Sorgfaltspflicht, Aufmerksamkeitsfehler, Erinnerungsfehler, rücksichtsloser Gebrauch oder nicht erfolgte Wartung [187, 344]

<sup>10</sup> Anwendungsfehler: Systematischer Fehler, bedingt durch Fehlkommunikation zw. Anwender und Gerät [187, 344]

Viertel der Fälle zu konstruktiven Änderungen für die weitere Verwendung der Medizinprodukte (23% für die Produkte im Feld und 27% bei Produkten in der Entwicklung) führte. Es kann davon ausgegangen werden, dass diese Studie nur die Spitze des Eisberges zeigt, da die meisten Anwender in deutschen Krankenhäusern nur geringe Kenntnisse über die gesetzliche Meldepflicht von Vorkommnissen gemäß der Medizinprodukte-Betreiberverordnung [275] resp. Sicherheitsverordnung [277] haben [274].

Nach Bleyer sind Röntgengeräte die am häufigsten von Bedienfehlern betroffenen Geräte, laut einer Studie von 209 Zwischenfällen (Abb. 20). Der prozentuale Anteil der Bedienfehler, gemessen an der Gesamtzahl aller Fehler, liegt bei 88,9%, gefolgt von den Hochfrequenz-Chirurgiegeräten mit einem Anteil von 87,8% und Infusionspumpen mit 66,7%. Bei Infusionspumpen sind jedoch die Folgen im Fehlerfall wesentlich schlimmer. In 33,3% der Fälle führte der Fehler zu einem tödlichen Zwischenfall, bei Hochfrequenz-Chirurgiegeräten „nur“ in 2% der Fälle und bei Röntgengeräten in keinem Fall [30].

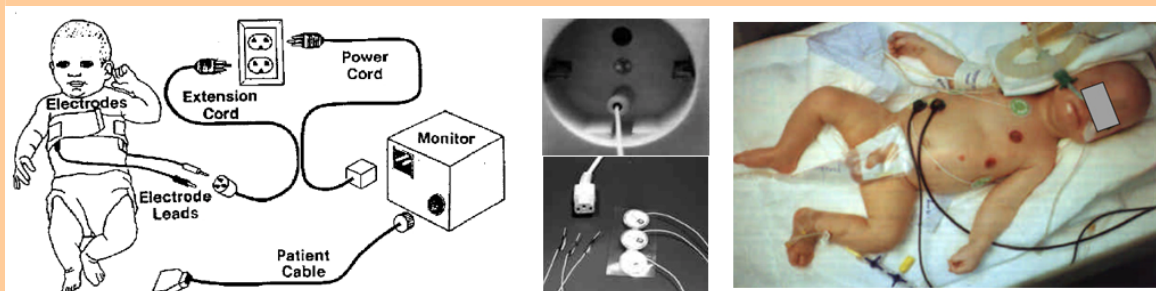


**Abb. 20: Fehlertypen unterschiedlicher Medizinprodukte**

Durch diese Ergebnisse erkennt man schnell, dass die Beachtung von ergonomischen Gesichtspunkten und Gebrauchstauglichkeitsanforderungen bei der Optimierung des medizinischen Arbeitsplatzes und der Verbesserung der verwendeten Geräte unverzichtbar ist. „Spitzenreiter“ der beteiligten Geräte bei fehlerverursachenden Situationen waren bei einer Untersuchung der FDA Blutglukosemessgeräte, Infusionspumpen und Hochfrequenz-Chirurgiegeräte [186]. Im Folgenden

werden zwei Szenarien vorgestellt, in denen die Benutzerschnittstelle zu Fehlern geführt hat.

1.) Abb. 21 zeigt, welche tragischen Auswirkungen die Verwechslung einer Steckverbindung an einem EKG-Gerät mit einer 220V Netzsteckbuchse hatte. Höchstwahrscheinlich hätten die Beachtung des Gefahrenpotentials und eine Analyse des Nutzungskontextes die potentielle Gefahrenquelle identifiziert und diese Fehler vermieden. Die Designer hätten ein Design entwickelt, welches diese Verwechslung bereits präventiv ausgeschlossen hätte.



**Abb. 21: Steckverbindung mit Risikopotential [352, 376]**

2.) Eine nicht verständliche Fehlermeldung des Linearbeschleunigers zur Anwendung in der Strahlentherapie „Therac-25“ der Firma Atomic Energy of Canada Limited war Ursache für drei tödliche Unfälle und weitere Schwerverletzte in den Jahren 1985-1987. Problematisch war, dass die Fehlermeldungen sehr kryptisch (z.B. Malfunktion 54) und nicht dokumentiert waren, sodass die Anwender nur wussten, dass ein Fehler aufgetreten war, aber nicht welcher. Durch eine einfache Bestätigungstaste konnte nach einer Fehlermeldung die Behandlung fortgesetzt werden. Da die Fehler sehr häufig (nicht selten 50× pro Tag) und meist ohne erkennbare Ursache auftraten, sank die Sensibilität der Anwender für Fehlermeldungen, und diese wurden nicht weiter beachtet. Dies führte dazu, dass Patienten tödlicher Strahlung ausgesetzt wurden [234].

Aber auch Einbußen auf Seiten der Anwender sind durch schlechte Ergonomie und Gebrauchstauglichkeit feststellbar. Grundsätzlich liegt der Fokus der medizinischen Behandlung auf dem Wohl und der Gesundheit des Patienten. Das medizinische Fachpersonal nimmt andererseits viele Unannehmlichkeiten stillschweigend hin. Stress, Zeit- und Erfolgsdruck beherrschen den medizinischen Alltag [16, 364]. Personalpolitik, personelle Engpässe und andauernd steigender Kostendruck erhöhen zusätzlich das Stresspotential für das medizinische Personal. Für Generalsekretär der deutschen Gesellschaft für Chirurgie (DGCH) ist dies mit fast 60% die Hauptursache von Unfällen und Gesundheitsrisiken im OP [16]. Die



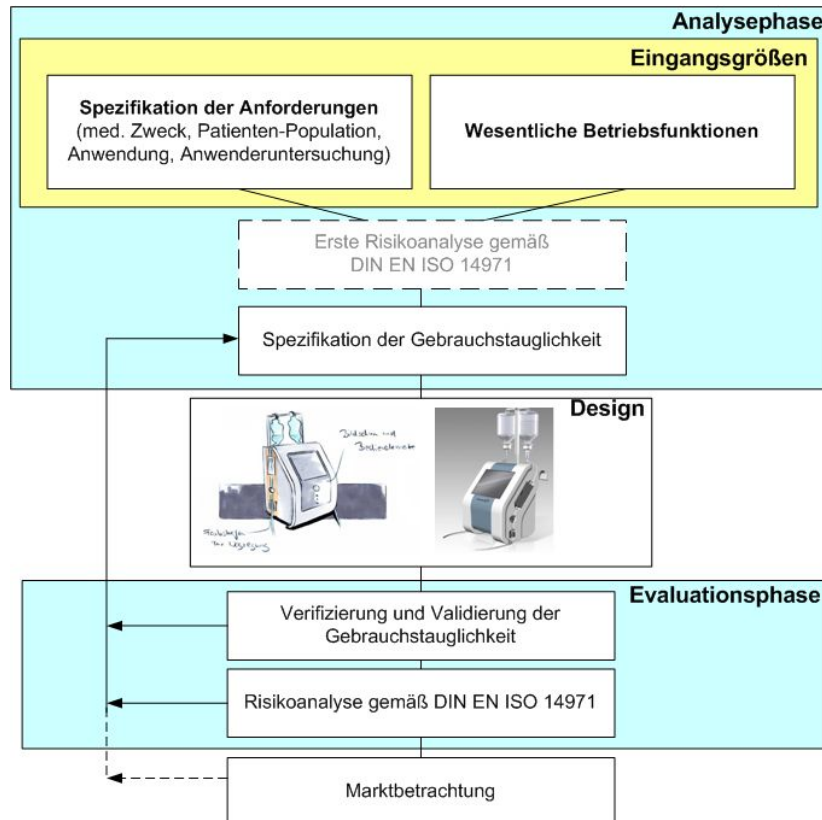
britische "Health and Safety Executive" (HSE) berichtete z.B. über eine erhöhte Häufigkeit von Erkrankungen bei Krankenschwestern infolge arbeitsbedingter Stress-Symptome [68]. Viele der Beschäftigten im OP müssen, nach eigenen Angaben, zumindest gelegentlich in unbequemer oder schmerzhafter Haltung arbeiten. Die Gründe hierfür sind Probleme mit den Geräten und Instrumenten, die Qualität der Beschaffenheit des OP-Mobiliars oder ungeeignete Arbeitsschritte und -abläufe [221]. Besonders häufig klagen die Anwender über Schmerzen im Rücken und Nacken, in den Schultern, den Oberarmen und den unteren Extremitäten. 35% der betroffenen Chirurgen, 59% der OP-Pflegekräfte und 44% der OP-Mitarbeiter haben sich wegen dieser Schmerzen bereits selbst behandelt oder haben sich in Behandlung begeben [221]. Oft sind derartige Probleme auch Ausgangspunkte für Gefährdungen bei Patienten oder beim OP-Personal.

### **2.2.3 Der Gebrauchstauglichkeitsentwicklungsprozess in der Medizin**

Der Gebrauchstauglichkeitsentwicklungsprozess in der Medizin unterscheidet sich generell kaum von dem in Kapitel 2.1.4 beschriebenen Prozess. Auch er sollte von den Bedürfnissen, Fähigkeiten, Eigenschaften und Zielen der Anwender getrieben sein und diese in den Mittelpunkt stellen. Da medizinische Geräte in der Regel sicherheitskritisch sind, muss dieser Prozess durch eine Sicherheitsbetrachtung und Elemente des Risikomanagements ergänzt werden. Nach Sawyer können demnach medizinische Geräte nur sicher und effizient verwendet werden, wenn der Hersteller während der Entwicklung die Wechselwirkungen zwischen der Arbeitsumgebung, den Nutzerfähigkeiten, dem Stresspotential und dem Gerätedesign beachtet [352].

Als zentrales Vorgehensmodell für einen auf Sicherheit bezogenen Gebrauchstauglichkeitsentwicklungsprozess für medizinische Geräte gilt der in der Norm DIN EN IEC 60601-1-6 beschriebene Prozess. Die Norm legt Anforderungen bzgl. eines vom Hersteller durchzuführenden Prozesses zur Analyse, Entwicklung, Gestaltung, Verifizierung und Validierung der Gebrauchstauglichkeit, bezogen auf die Sicherheit von Medizinprodukten, fest [97]. Dieser Ergonomieprozess schätzt Risiken ein und verringert Risiken, die von Problemen mit der Gebrauchstauglichkeit sowohl bei bestimmungsgemäßem Gebrauch als auch bei Benutzungsfehlern ausgelöst werden. Entsprechend der Norm ist er lediglich auf sicherheitsbezogene Gebrauchstauglichkeit ausgerichtet, kann jedoch ohne weiteres auch auf eine allgemeine Gebrauchstauglichkeit übertragen werden.

Ergänzend zu dem in Kapitel 2.1.4 beschriebenen iterativen Prozess sieht der Entwicklungsprozess für Gebrauchstauglichkeit medizinischer Produkte (Abb. 22) eine spezielle Risikoanalyse für Gebrauchstauglichkeitsprobleme gemäß DIN EN ISO 14971 vor, welche während der Entwicklung fortwährend im Zuge des iterativen Prozesses wiederholt werden soll.



**Abb. 22: Entwicklungsprozess für gebrauchstaugliche medizinische Produkte**

Im medizinischen Arbeitssystem hilft dieser Entwicklungsprozess, die Verbesserung der menschlichen Leistungsfähigkeit und die Verringerung der Risiken im Zusammenhang mit Nutzungsfehlern zu fördern. Ziele von gebrauchstauglich gestalteten Produkten sind für den **Patienten** in erster Linie eine sichere und qualitativ hochwertige schnelle Diagnose und Behandlung, aber auch die Versorgung durch leistungsfähiges und zufriedenes Personal. Für den **Anwender** stehen die Leichtigkeit, Flexibilität und Genauigkeit der Anwendung durch gebrauchstaugliche Produkte im Vordergrund. Durch effiziente und intuitive Prozesse würden Anwender zufriedengestellt und hätten mehr Zeit für den Patienten, da dieselbe Arbeit in kürzerer Zeit erledigt werden könnte. Anwender stünden folglich weniger unter Stress. **Betreiber** würden einerseits durch gesteigerte Produktivität und Effizienz und eine damit verbundene Kostenersparnis profitie-

ren, andererseits durch erhöhte Sicherheit, ein zufriedenes Personal und einer dadurch geringeren Personalfuktuation. Auch wirkt sich eine hohe Patientenzufriedenheit positiv für Betreiber aus. **Hersteller** profitieren durch Abgrenzung von den Mitbewerbern und durch erhöhte Marktbehauptung, dank zufriedener Kunden, die ihre Produkte als sicher und an ihre Bedürfnisse angepasst ansehen, sowie durch geringere Support- und Entwicklungskosten [187]. Abb. 23 fasst die durch diesen Prozess entstehenden Vorteile für Anwender, Betreiber und Hersteller medizinischer Produkte zusammen.

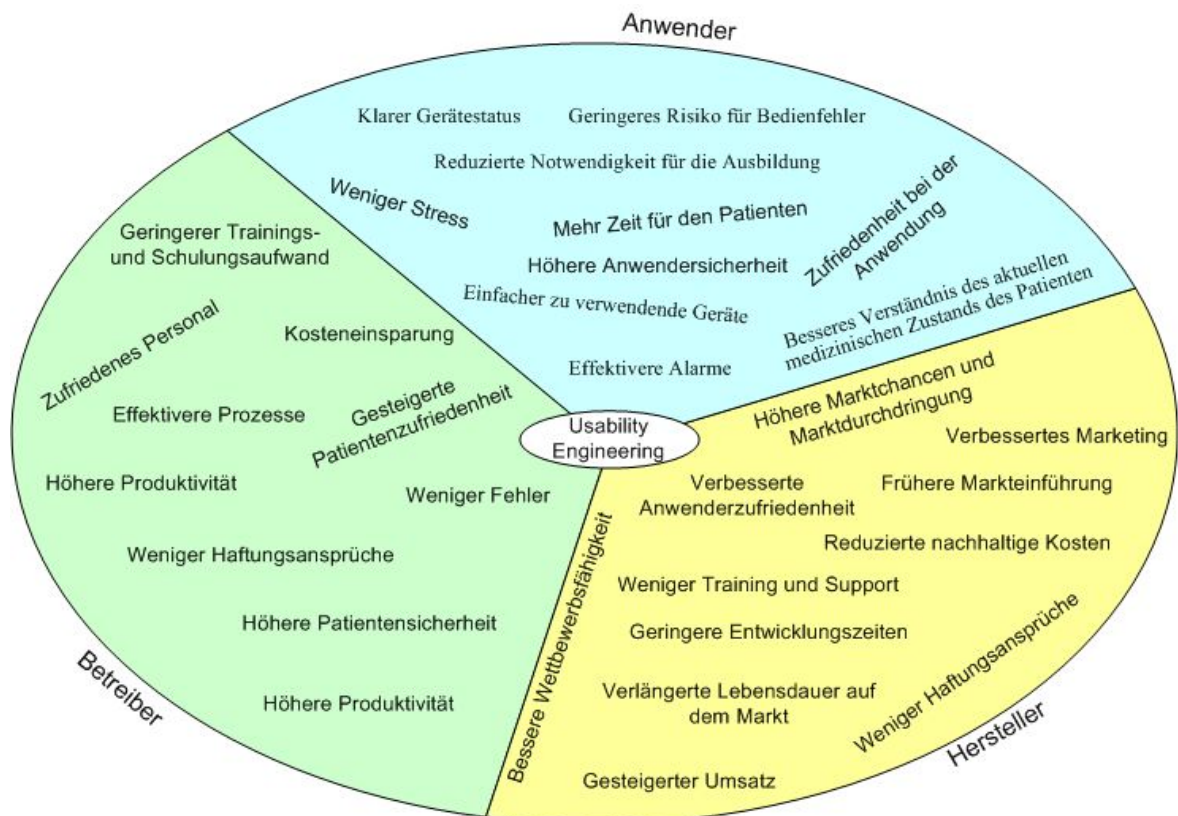


Abb. 23: Vorteile eines benutzerzentrierten Entwicklungsprozesses in der Medizin

## 2.3 Gesetze, Normen und Richtlinien

Im Folgenden werden die für diese Arbeiten wichtigen Gesetze, Normen und Richtlinien vorgestellt.

### 2.3.1 Gesetze

Dass Medizinprodukte alle einschlägigen Sicherheitsanforderungen erfüllen, bestätigen Hersteller durch eine schriftliche EG-Konformitätserklärung, deren Bestandteil neben einer Produktbeschreibung und dem Namen des Herstellers

auch eine Auflistung der eingehaltenen Normen und technischen Spezifikationen ist. In der EG-Konformitätserklärung wird üblicherweise darüber hinaus beschrieben, welche anderen zutreffenden EG-Richtlinien eingehalten sind, oder ob dafür ggf. geltende Übergangsregelungen in Anspruch genommen werden [18]. Diese Einhaltung wird nach außen durch das CE-Kennzeichen signalisiert, welches für Produkte in Zusammenhang mit der Produktsicherheit steht. Durch ein CE-Kennzeichen signalisiert der Hersteller, dass das Produkt geltenden EU-Richtlinien entspricht und ein Konformitätsbewertungsverfahren durchgeführt wurde, welches als Voraussetzung für die Inverkehrgabe und den Vertrieb von Medizinprodukten gilt. Die Bewertung kann „im Auftrag des Herstellers“ auch von einem Bevollmächtigten einer sog. „Benannten Stelle“ im Sinne des § 3 Nr. 16 MPG durchgeführt werden (vgl. Kap. 1.2.2.3). Deren Einbindung in das Konformitätsbewertungsverfahren wird durch die zusätzliche Platzierung einer vierstelligen Nummer neben dem CE-Kennzeichen signalisiert.

Gemäß Art. 95 des EG-Vertrages<sup>11</sup> werden für zahlreiche Produkte Sicherheits- und Gesundheitsanforderungen festgelegt, die nicht unterschritten werden dürfen. Für Medizinprodukte legt auf europäischer Ebene die Medizinprodukterichtlinie (93/42/EWG), resp. die EU-Richtlinie 98/79/EWG, für In-vitro-Diagnostik und für aktive implantierbare medizinische Geräte die EU-Richtlinie 90/385/EWG die Voraussetzungen für medizinische Geräte fest. Sie fordern die CE-Kennzeichnung sowie in einigen Fällen eine zusätzliche Begutachtung durch Prüfstellen [131, 132, 133].

Die Richtlinie 93/42/EWG fordert:

*„Die Produkte müssen so ausgelegt und hergestellt sein, dass ihre Anwendung unter den vorgesehenen Bedingungen und zu den vorgesehenen Zwecken weder den klinischen Zustand und die Sicherheit der Patienten noch die Sicherheit und die Gesundheit der Anwender oder gegebenenfalls Dritter gefährdet, wobei etwaige Risiken im Zusammenhang mit der vorgesehenen Anwendung gemessen am Nutzen für den Patienten vertretbar und mit einem hohen Maß an Gesundheitsschutz und Sicherheit vereinbar sein müssen.*

---

<sup>11</sup> Konsolidierte Fassung des Vertrags über die Europäische Union, Amtsblatt Nr. C 325 vom 24.12.2002

*Dazu gehört*

- eine weitestgehende Verringerung der durch Anwendungsfehler bedingten Risiken aufgrund der ergonomischen Merkmale des Produkts und der Umgebungsbedingungen, in denen das Produkt eingesetzt werden soll (Produktauslegung im Hinblick auf die Sicherheit des Patienten), sowie
- die Berücksichtigung der technischen Kenntnisse, der Erfahrung, Aus- und Weiterbildung sowie gegebenenfalls der medizinischen und physischen Voraussetzungen der vorgesehenen Anwender (Produktauslegung für Laien, Fachleute, Behinderte oder sonstige Anwender).“

Auch durch die Einführung der Zusatznorm DIN EN IEC 60601-1-6 [97] müssen seit 2006 Hersteller medizinisch-elektrischer Geräte und seit 2008 durch die Norm DIN EN IEC 62366 Hersteller aller medizinischer Geräte die Gebrauchstauglichkeit bei ihren Geräten nachweisen, um für sie eine CE-Kennzeichnung führen zu dürfen [362]. Im Zuge der Umsetzung europäischer Rechtsvorschriften in nationales Recht wurde in Deutschland die EU-Richtlinie in Form des Medizinproduktegesetzes übertragen [276]. Zusätzlich existieren folgende Rechtsgrundlagen, welche im Kontext dieser Arbeit von Interesse sind<sup>12</sup>:

- Medizinprodukteverordnung (MPV)
- Medizinprodukte-Betreiberverordnung (MPBetreibV)
- Medizinprodukte-Sicherheitsplanverordnung (MPSV)
- Informationssystem über Medizinprodukte (DIMDIV)

#### MPG – Medizinproduktegesetz

Das Medizinproduktegesetz<sup>13</sup> (MPG) löste am 1. Januar 1995 die bis dahin bestehende Medizingeräteverordnung ab. Es wurde am 7. August 2002 neu gefasst und ist die nationale Umsetzung des europäischen Rechts in deutsches Recht. Es regelt den Verkehr von Medizinprodukten und sorgt für deren Sicherheit, Eignung und Leistung sowie für die Gesundheit und den erforderlichen Schutz der Patienten, der Anwender und Dritter [276]. Es enthält neben technischen und medizinischen Anforderungen auch Informationsanforderungen und Vorschriften für Betreiber und Anwender medizinischer Produkte. Als Voraussetzung für deren Vertrieb von Medizinprodukten gilt die in §7 geforderte CE-Kennzeichnung, die

<sup>12</sup> Gesetzestexte unter: <http://www.dimdi.de/static/de/mpg/recht/index.htm> (03.12.2009)

<sup>13</sup> Gesetzestext unter: <http://bundesrecht.juris.de/mpg/index.html> (03.12.2009)

ihrerseits eine Übereinstimmung nach §8 MPG mit den harmonisierenden Normen voraussetzt.

Aus der in §3 MPG definierten Zweckbestimmung wird deutlich, dass sich der anwendende Arzt bei einem Medizinprodukt darauf verlassen kann, dass innerhalb der in der Produktdokumentation angegebenen Zweckbestimmung die Behandlung des Patienten in ausreichender Sicherheit erfolgt. Auf der Basis des MPG kann sich der Arzt außerdem unabhängig von seinem technischen Verständnis auf die Wirksamkeit und Funktionalität des Produktes verlassen. Die Hersteller müssen innerhalb der Zweckbestimmung die Sicherheit gewährleisten [121]. Sie müssen Nutzen-Risikoanalysen durchführen und durch geeignete Maßnahmen Schäden und Risiken vermeiden bzw. beschränken. In §4 MPG wird dem Hersteller unter Strafe verboten, ein nicht ausreichend sicheres Medizinprodukt im Markt einzuführen. Außerhalb der vom Hersteller ausgewiesenen Zweckbestimmung handelt der Arzt eigenverantwortlich. Weiterhin dürfen Medizinprodukte nicht betrieben werden, wenn der Verdacht besteht, dass sie die Sicherheit und Gesundheit von Personen, selbst bei sachgemäßer Anwendung, Instandhaltung und Einhaltung der Zweckbestimmung, gefährden. Bei Nichteinhaltung verstoßen sowohl Hersteller als auch Betreiber gegen das MPG. Erlangt ein Krankenhaus (Betreiber) davon Kenntnis, dass ein Produkt gesundheitsgefährdende Gebrauchstauglichkeitsmängel hat, müsste somit der Betrieb desselben strikt untersagt werden. Andernfalls handeln die Verantwortlichen gegen Richtlinien und Gesetze. Zwar enthält das MPG keine speziellen Haftungsregelungen, jedoch greifen hierbei maßgeblich das Produkthaftungsgesetz und die Haftungsregelung des §823 BGB. Sanktionen reichen von Ordnungswidrigkeits- bis hin zu Strafverfahren. Die Rechtsanwälte der Kanzlei „Nörr, Stiefenhofer, Lutz“ berichten von strafrechtlicher Verfolgung bei Körperverletzung nach §§ 223, 229 StGB, bei Totschlag oder fahrlässiger Tötung nach §§ 212, 222 StGB, mit dem Tatvorwurf: Inverkehrgabe eines fehlerhaften Medizinproduktes oder fehlerhafte Anwendung des Medizinproduktes [377].

Insgesamt stellt das MPG hohe Anforderungen an die Sicherheit von Medizinprodukten. Explizite Gebrauchstauglichkeitsanforderungen werden zwar nicht gestellt, jedoch wird durch die in §8 geforderte Normkonformität dem Hersteller medizinischer Geräte die Beachtung der Gebrauchstauglichkeit entsprechend der Zusatznormen DIN EN IEC 60601-1-6, DIN EN IEC 62366 und DIN EN ISO 14971 auferlegt, welche als harmonisierende Normen angesehen werden können.

### MPBetreibV - Medizinprodukte-Betreiberverordnung

Die Medizinprodukte-Betreiberverordnung<sup>14</sup> (MPBetreibV) in ihrer Neufassung vom 21.08.2002 ist gültig für „das Errichten, Betreiben, Anwenden und Instandhalten von Medizinprodukten nach § 3 MPG mit Ausnahme der Medizinprodukte zur klinischen Prüfung oder zur Leistungsbewertungsprüfung“ (§1) und stellt somit ein bedeutsames Regelwerk für die Krankenhauspraxis dar.

Nach dieser Verordnung dürfen Medizinprodukte nur entsprechend ihrer Zweckbestimmung und nach den allgemein anerkannten Regeln der Technik sowie den Arbeitsschutz- und Unfallverhütungsvorschriften errichtet, betrieben, angewendet und instand gehalten werden (§2 Abs. 1). Der hierfür berechnigte Personenkreis wird gemäß §2 auf Personen begrenzt, die eine dafür erforderliche Ausbildung oder Kenntnis und Erfahrung besitzen (Abs.2), sich von der Funktionsfähigkeit und dem ordnungsgemäßen Zustand des Medizinproduktes überzeugt und die Gebrauchsanweisung sowie die sonstigen beigefügten, sicherheitsbezogenen Informationen und Instandhaltungshinweise beachtet haben (Abs. 5). Zusätzlich fordert die Verordnung bei aktiven Medizingeräten gemäß §5 eine, von einer durch den Betreiber beauftragten Person, durchzuführende Einweisung. Des Weiteren dürfen Medizinprodukte nach §2 Abs.3 nur dann betrieben und angewendet werden, wenn sie unter Berücksichtigung der Zweckbestimmung und der Sicherheit der Patienten, Anwender, Beschäftigten oder Dritten dazu geeignet sind.

### MPSV - Medizinprodukte-Sicherheitsplanverordnung

Die Medizinprodukte-Sicherheitsplanverordnung<sup>15</sup> (MPSV) regelt seit dem 28.06.2002 die Verfahren zur Erfassung, Bewertung und Abwehr von Risiken im Verkehr oder in Betrieb befindlicher Medizinprodukte. In §3 verpflichtet sie sowohl Verantwortliche nach §5 MPG (Satz 1) sowie berufliche und gewerbliche Betreiber (Satz 2), aufgetretene Vorkommnisse und ggf. Rückrufe der zuständigen Bundesoberbehörde zu melden. Diese kann daraufhin eine Risikobewertung vornehmen (§8), um „festzustellen, ob ein unvertretbares Risiko vorliegt und welche korrektiven Maßnahmen geboten sind“ (§8). Darüber hinaus kann sie bei nicht ausreichender Risikominderung nach den §§ 14 und 15 die notwendigen Maßnahmen ergreifen, „um das Betreiben oder Anwenden der betroffenen Medi-

---

<sup>14</sup> Abruf unter: <http://bundesrecht.juris.de/mpbetreibv/> (03.12.2009)

<sup>15</sup> Abruf unter: <http://www.buzer.de/gesetz/6622/index.htm> (03.12.2009)

zinprodukte zu untersagen oder einzuschränken“ (§17). Zu diesen Risiken gehören auch Fehler welche durch Gebrauchstauglichkeitsprobleme entstanden sind.

#### MPV - Medizinprodukteverordnung

Die am 01.01.2002 in Kraft getretene Medizinprodukteverordnung<sup>16</sup> (MPV) regelt „die Bewertung und Feststellung der Übereinstimmung von Medizinprodukten mit den grundlegenden Anforderungen gemäß § 7 des Medizinproduktegesetzes (Konformitätsbewertung) [...]“. Unter Verweis auf bestehende EU Richtlinien regelt die Verordnung in §3 die allgemeinen Vorschriften zur Durchführung der Konformitätsbewertung. Hierbei sind nach Satz 3 den Benannten Stellen alle Informationen, welche für die Bewertung und Durchführung einer Überprüfung notwendig sind, bereitzustellen.

#### ArbSchG - Arbeitsschutzgesetz

Die Einhaltung der Vorschriften des Arbeitsschutzes wird explizit von der MPBetreibV in §2 gefordert. Das Arbeitsschutzgesetz<sup>17</sup> (ArbSchG), dient seit dem 21. August 1996 dazu, „Sicherheit und Gesundheitsschutz der Beschäftigten bei der Arbeit durch Maßnahmen des Arbeitsschutzes zu sichern und zu verbessern. Es gilt in allen Tätigkeitsbereichen“ (§1. Abs.1) und legt somit auch im medizinischen Bereich dem Arbeitgeber die Verpflichtung auf, erforderliche Maßnahmen des Arbeitsschutzes durchzuführen, welche die Sicherheit und Gesundheit der Beschäftigten bei der Arbeit beeinflussen, und eine Verbesserung von Sicherheit und Gesundheitsschutz anzustreben (§3). Hierzu muss er die Gefährdungen laut §5 ermitteln und beurteilen. Das Gesetz betont, dass sich derartige Gefährdungen besonders durch die Gestaltung und die Einrichtung der Arbeitsstätte und des Arbeitsplatzes, die Gestaltung, die Auswahl und den Einsatz von Arbeitsmitteln, resp. von Arbeitsstoffen, Maschinen, Geräten und Anlagen und durch den Umgang mit ihnen sowie durch die Gestaltung von Arbeits- und Fertigungsverfahren, Arbeitsabläufen und Arbeitszeit und deren Zusammenwirken ergeben.

Die Beschäftigten haben die Pflicht „nach ihren Möglichkeiten [...] für ihre Sicherheit und Gesundheit bei der Arbeit Sorge zu tragen [...]“ und „[...] auch für die Sicherheit und Gesundheit der Personen zu sorgen, die von ihren Handlungen

---

<sup>16</sup> Abruf unter: <http://www.buzer.de/gesetz/226/index.htm> (03.12.2009)

<sup>17</sup> Abruf unter: <http://www.gesetze-im-internet.de/arbschg/> (03.12.2009)



oder Unterlassungen bei der Arbeit betroffen sind.“ Im medizinischen Umfeld gehören dazu auch Personal und Patienten.

Entsprechend dem ArbSchG sind somit Arbeitgeber und Beschäftigte dazu verpflichtet, „[...] Maßnahmen zur Vermeidung von Unfällen bei der Arbeit und arbeitsbedingten Gesundheitsgefahren einschließlich Maßnahmen der menschengerechten Gestaltung der Arbeit“ (§ 2 Abs. 1) zu treffen. Sicher und fehlerfrei zu bedienende Geräte zählen hierzu. Das Arbeitsschutzgesetz verweist hierbei allgemein auf mögliche Gefahrenquellen. Für spezielle Bereiche werden sie beispielsweise durch die Bildschirmarbeitsverordnung (BildschArbV) oder die Arbeitsstättenverordnung (ArbStättV) konkretisiert.

#### Bildschirmarbeitsverordnung (BildscharbV)

Die Bildschirmarbeitsverordnung<sup>18</sup> (BildscharbV) ist eine Verordnung speziell für die Sicherheit und den Gesundheitsschutz bei der Arbeit an Bildschirmgeräten. Obwohl diese Verordnung lediglich für bildschirmbasierte medizinische Geräte zur Anwendung kommt (§1), sei sie dennoch an dieser Stelle erwähnt, da sie explizit im Anhang unter der Überschrift „Zusammenwirken Mensch-Arbeitsmittel“ eine Anwendung von Grundsätzen der Ergonomie und Benutzerfreundlichkeit fordert. Durch diese Verordnung wurden die Ansprüche von Arbeitnehmern auf benutzerfreundliche Benutzungsschnittstellen gesetzlich fixiert [27].

#### Sozialgesetzbuch (SGB)

Das Sozialgesetzbuch<sup>19</sup> (SGB) regelt die wesentlichen Bereiche des Sozialrechts. Das fünfte Buch des Sozialgesetzbuchs (Gesetzliche Krankenversicherung) betrifft Organisation, Versicherungspflicht und Leistungen der gesetzlichen Krankenkassen sowie deren Rechtsbeziehungen zu weiteren Leistungserbringern (Ärzte, Zahnärzte, Apotheker etc.). Durch §135a verpflichtet es die Leistungserbringer nach dem jeweiligen Stand der wissenschaftlichen Erkenntnisse „zur Sicherung und Weiterentwicklung der Qualität der von ihnen erbrachten Leistung[...]“. Weiterhin verpflichten sich „Vertragsärzte, medizinische Versorgungszentren, zugelassene Krankenhäuser [...] 1. sich an einrichtungsübergreifenden Maßnahmen der Qualitätssicherung zu beteiligen, die insbesondere zum Ziel haben, die

---

<sup>18</sup> Abruf unter: [http://bundesrecht.juris.de/bildscharbv/\\_\\_1.html](http://bundesrecht.juris.de/bildscharbv/__1.html) (03.12.2009)

<sup>19</sup> Abruf unter: [http://bundesrecht.juris.de/sgb\\_5/index.html](http://bundesrecht.juris.de/sgb_5/index.html) (03.12.2009)

Ergebnisqualität zu verbessern und 2. einrichtungsintern ein Qualitätsmanagement einzuführen und weiterzuentwickeln“ (SGB 135a).

#### Geräte- und Produktsicherheitsgesetz GPSG

Das Geräte- und Produktsicherheitsgesetz<sup>20</sup> (GPSG) regelt seit dem 1. Mai 2004 das „Inverkehrbringen und Ausstellen von Produkten, das selbständig im Rahmen einer wirtschaftlichen Unternehmung erfolgt“ und „die Errichtung und den Betrieb überwachungsbedürftiger Anlagen, die gewerblichen oder wirtschaftlichen Zwecken dienen oder durch die Beschäftigte gefährdet werden können“. Nach dem Gesetz darf ein Produkt nur in den Verkehr gebracht werden, wenn es so beschaffen ist, dass bei bestimmungsgemäßer Verwendung oder vorhersehbarer Fehlanwendung Sicherheit und Gesundheit von Verwendern oder Dritten nicht gefährdet werden (§4).

#### Regeln der FDA für die USA

Die in den USA ansässige Food and Drug Administration (FDA) hat eigens für das Thema Gebrauchstauglichkeit in der Medizin ein Human Factors Programm<sup>21</sup> eingeführt und fordert im Teil 21 des „Code of Federal Regulations<sup>22</sup>“ im §820.30 Teile c, f und g, dass auch bei medizinischen Geräten Aspekte der Gebrauchstauglichkeit angemessen berücksichtigt werden müssen. Teil c legt fest, dass jeder Hersteller angemessene Verfahren verwenden muss, die sicherstellen, dass die Designanforderungen an das Gerät geeignet sind und welche die beabsichtigte Verwendung und die Bedürfnisse der Anwender und Patienten berücksichtigen. In Teil f fordert die FDA, dass jeder Hersteller Verfahren verwendet, welche die Designentscheidung überprüfen. Teil g besagt, dass die entwickelten Geräte unter Einbeziehung der Nutzerbedürfnisse und der beabsichtigten Verwendung unter tatsächlichen oder simulierten Bedingungen validiert werden müssen [397]. Zur Unterstützung der Hersteller bei diesem Prozess hat die FDA gerätespezifische und allgemeine Leitlinien (engl.: guidance) herausgegeben, welche konkrete Empfehlungen für die Benutzerschnittstelle medizinischer Geräte enthalten. Weiterhin veröffentlicht sie offizielle, durch beispielsweise die AAMI (Association for the Advancement of Medical Instrumentation) und die IEC (International Electro-

---

<sup>20</sup> Abruf unter: <http://bundesrecht.juris.de/gpsg/> (03.12.2009)

<sup>21</sup> Abruf unter: <http://www.fda.gov/cdrh/humanfactors/index.html> (03.12.2009)

<sup>22</sup> Abruf unter: <http://www.access.gpo.gov/cgi-bin/cfrassemble.cgi?title=199821> (03.12.2009)

technical Commission) anerkannte Normen, welche auf der Internetseite der FDA abrufbar sind.

### 2.3.2 Normen

Deutsche Rechtsprechung und EU-Richtlinien definieren die grundlegenden Anforderungen an die Sicherheit von Medizinprodukten. Die technische Konkretisierung erfolgt in „harmonisierenden“ Normen. Normen sind zunächst freiwillige Übereinkünfte und erreichen erst über die Bezugnahme in Gesetzen, Verordnungen oder Verträgen verbindlichen Status. Sie stellen den Stand der Technik und der gesicherten arbeitswissenschaftlichen Erkenntnisse dar, die sich im Laufe der Zeit verändern können [151]. Der Zweck einer Norm entscheidet über ihren Inhalt. Petrick & Reihens unterscheiden Verständigungsnormen<sup>23</sup>, Verfahrensnormen<sup>24</sup> und Produktnormen<sup>25</sup> [310].

Im Bereich der Ergonomie und der Mensch-Maschine-Kommunikation hat sich in Deutschland frühzeitig der Normungsausschuss Informationsverarbeitung des DIN (Deutsches Institut für Normung e.V.) hervorgetan. Bereits in den 70er Jahren hat er eine Vorreiterrolle im Bereich der Softwareergonomie eingenommen [196]. Auf internationaler Ebene ist die ISO (International Organization for Standardization) zu nennen, welche die ergonomischen Hard- und Softwareanforderungen für die in diesem Bereich besonders wichtige Norm 9241 formuliert hat. Die DIN EN ISO 9241 beschreibt die Interaktion zwischen Mensch und Computer. 2006 wurde die bisherige Einschränkung auf die „Ergonomische(n) Anforderung(en) für Bürotätigkeiten mit Bildschirmgeräten“ durch die Namensänderung auf „Ergonomie der Mensch-System-Interaktion“ aufgehoben. Im Bereich der Gebrauchstauglichkeit ist Teil 11 von großer Bedeutung, da der Begriff in diesem Teil definiert wird. Den Entwicklungsprozess für Gebrauchstauglichkeit in der Medizin beschreiben die Ergänzungsnorm DIN EN IEC 60601-1-6 „Medizinische elektrische Geräte: Allgemeine Festlegungen für die Sicherheit – Ergänzungs-

---

<sup>23</sup> Festlegungen zu Begriffen, Zeichen, Systemen

<sup>24</sup> Festlegungen zu Abläufen, Merkmalen bzw. Merkmalswerten für Tätigkeiten/ Verfahren/ Prozessen zur Anwendung in der Planung, Realisierung und Nutzung von Einheiten oder zu deren Prüfung

<sup>25</sup> Festlegungen zu Merkmalen bzw. Merkmalswerten für materielle oder immaterielle Produkte (also auch Dienstleistungen) in Form von Sortenbeschreibungen und von Beschaffenheitsanforderungen, insbesondere anhand von Qualitätsanforderungen. Dazu gehören auch spezielle Qualitätsmerkmale wie Kompatibilität, Sicherheit, Instandhaltbarkeit, Umweltschutz usw.

norm: Gebrauchstauglichkeit“ [97] und die Norm DIN EN IEC 62366 „Medizinprodukte - Anwendung der Gebrauchstauglichkeit auf Medizinprodukte“.

Für spezielle Anforderungen werden verschiedene Normen benötigt. Im Folgenden werden daher nur die wichtigsten Normen für Gebrauchstauglichkeit im medizinischen Kontext genau beschrieben. Viele weitere Normen sind für spezielle Anwendungsfälle erstellt worden.

### **Wichtige Normen bei der Gebrauchstauglichkeitsbetrachtung in der Medizin**

Die internationale Norm ISO 14971 in der deutschen Fassung DIN EN ISO 14971 „Medizinprodukte - Anwendung des Risikomanagements auf Medizinprodukte“, in der aktuellen Version von 2007, legt „einen Prozess für einen Medizinproduktehersteller [...] zur Identifizierung der mit Medizinprodukten verbundenen Gefährdungen [...]“ fest. Dieser dient „der Einschätzung und der Bewertung zugehöriger Risiken zur Beherrschung dieser Risiken und der Überwachung der Wirksamkeit von Maßnahmen zur Risikobeherrschung“ [108].

Die Norm ist eng an die produktbezogenen Prozesse angegliedert und kann in allen Phasen der Entwicklung eingesetzt werden. Sie dient dem Schutz von Anwendern, Patienten und Dritten. In Anhang C der Norm werden auch Gebrauchstauglichkeitsbestandteile in die Risikobetrachtung mit einbezogen, jedoch werden keine expliziten Maßnahmen zur Aufdeckung von Risiken durch Gebrauchstauglichkeitsdefizite vorgeschlagen. Deshalb könnten bei alleiniger Anwendung dieser Norm Gebrauchstauglichkeitsdefizite aufgrund ihrer schweren Erfassbarkeit schnell verborgen bleiben.

Dieser Mangel wird durch die Grundlagennorm für Sicherheit medizinischer elektrischer Geräte DIN EN IEC 60601-1 "Medizinische elektrische Geräte; Teil 1: Allgemeine Festlegungen für die Sicherheit" in Verbindung mit der Ergänzungsnorm: „Gebrauchstauglichkeit“ [97] behoben. Die DIN EN IEC 60601-1 ist Grundlage für die Sicherheit von medizinisch-elektrischen Geräten bei der Entwicklung, Produktion und Anwendung. Der Geltungsbereich überschreitet die Schwelle der Krankenhausumgebung und gilt auch für medizinische Geräte/Systeme im Hausgebrauch. Die Norm fordert eine konsequente Einführung des Risikomanagements nach DIN EN ISO 14971 bei der Beurteilung von Gefährdungen oder möglichen Gefährdungen.

Die Ergänzungsnorm 6 der DIN EN IEC 60601-1 legt allen Herstellern, die ein nach 93/42/EWG definiertes elektrisch-medizinisches Gerät in Verkehr bringen

wollen, eine Ergonomiebetrachtung bei der Entwicklung dieser Gerätes fest. Sie beschreibt „Anforderungen an einen Prozess für die Analyse, die Konzeption, die Verifizierung und Validierung der Gebrauchstauglichkeit in Bezug auf die Sicherheit von medizinischen elektrischen Geräten“, welche deren Gebrauchstauglichkeit durch entsprechende Anforderungen und einen direkten Bezug zu dem in DIN EN ISO 14971 festgelegten Risikomanagementprozess sicherstellen sollen [97]. Die hierfür notwendigen Aktivitäten im Zusammenhang mit der Konformitätsprüfung liegen im Verantwortungsbereich des Herstellers. Die sich aus der Norm DIN EN IEC 60601-1-6 ergebenden Aufgaben sind durch ihn in geeigneter Weise in den Entwicklungsprozess zu integrieren. Der Anwendungsbereich dieser Norm geht über den „bestimmungsgemäßen Gebrauch“ (entspr. der vom Hersteller definierten Zweckbestimmung (MPG §3 Abs. 10)) hinaus und schließt eine „vernünftigerweise vorhersehbare Fehlbenutzung“ mit ein [97].

Die Norm wird als ein „minimum safety standard“ angesehen, der ein Mindestmaß an Sicherheit sicherstellen soll. Das bedeutet, dass lediglich die Sicherheit durch die Gebrauchstauglichkeit gewährleistet sein muss. Eine angenehme und einfache Bedienung wird durch diese Norm nicht garantiert, solange sie kein Sicherheitsrisiko darstellt.

Die Norm gilt als harmonisierende Norm und wird im Rahmen der Konformitätsbewertungsverfahren nach der EU-Richtlinie für medizinische Produkte 93/42/EWG und der Umsetzung in nationales Recht durch das MPG angewandt. Hersteller müssen, im Sinne der ihnen auferlegten Nachweisführung, zur Erfüllung dieser Norm eine „Akte der Entwicklung für Gebrauchstauglichkeit“, auch Ergonomieakte genannt, anlegen, in der die Aktivitäten zur Entwicklung für Gebrauchstauglichkeit aufgezeichnet werden, und anhand derer Zertifizierungsstellen die normgerechte Entwicklung des Medizinproduktes überprüfen. Verpflichtende Inhalte der Ergonomieakte sind:

- Dokumente, wie z.B. die Produktakte oder die Risikomanagement-Akte,
- Überlegungen bzgl. Begleitpapieren und Kennzeichnungen auf dem Gerät,
- Nachweis über den iterativen Entwicklungsprozess für Gebrauchstauglichkeit,
- Ergebnis des Entwicklungsprozesses der Gebrauchstauglichkeit,
- Spezifikation der Anwendung des Gerätes,
- Wesentliche Betriebsfunktionen,

- Risikoanalyse,
- Spezifikation der Gebrauchstauglichkeit,
- Verifizierung der Gebrauchstauglichkeit,
- Validierungsplan für die Gebrauchstauglichkeit,
- Validierung der Gebrauchstauglichkeit.

Für Geräte, die sich bereits auf dem Markt befinden, muss der Hersteller nicht nachträglich eine derartige Dokumentation erstellen. Lediglich ein Nachweis durch eine GAP-Analyse<sup>26</sup> muss erbracht werden, dass die, durch die DIN EN IEC 60601-1-6 entstandenen Anforderungen keine Auswirkungen auf das Gerät haben. Als Basis für die GAP-Analyse können laut dem Deutschen Industrieverband für optische, medizinische und mechatronische Technologien e.V. (SPECTARIS) unterschiedliche Quellen herangezogen werden, wie z.B. Rückmeldungen aus dem eigenen Beobachtungs- und Meldesystem, Meldungen des BfArM oder anderer Behörden wie der FDA oder auch Informationen zu Konkurrenzprodukten [97, 425]. Bei negativem Ergebnis der GAP-Analyse, d.h. bei Verbesserungsfähigkeit der Gebrauchstauglichkeit, muss bei den anfallenden Änderungen der in der DIN EN IEC 60601-1-6 geforderte Prozess eingehalten werden.

Die Einschränkung der DIN EN IEC 60601-1-6 auf elektrisch-medizinische Geräte wurde von der 2008 harmonisierten Norm DIN EN IEC 62366 aufgehoben. Diese Norm überträgt die Grundgedanken der DIN EN IEC 60601-1-6 auf alle Medizinprodukte. Die Norm stellt nahezu gleiche Anforderungen an den Prozess zur Analyse, Entwicklung, Gestaltung, Verifizierung und Validierung der Gebrauchstauglichkeit, bezogen auf die Sicherheit von Medizinprodukten, und an dessen Dokumentation wie die DIN EN IEC 60601-1-6. Sie hat das Potential, die DIN 60601-1-6 zu ersetzen, da sie deren Einschränkung auf medizinisch-elektrische Geräte aufhebt und für alle medizinischen Geräte gilt.

### **Normen der Ergonomie**

Normen aus dem Bereich der Ergonomie waren lange hauptsächlich auf Hardware bezogen. Infolge weiterer Durchdringung von Computertechnologien wurden seit Anfang der 80er Jahre ergonomische Normen für Benutzerschnittstellen definiert. Im Gegensatz zu den Hardwarenormen sorgen sie nicht dafür, dass die

---

<sup>26</sup> GAP-Analyse (Lückenanalyse): Analyse, durch die sich Abweichungen von einem geplanten Ziel feststellen lassen

Benutzerschnittstellen gleich gestaltet sind (z.B. einheitliche Zweipol-Stecker, DIN A4), sondern geben Herstellern vielmehr wertvolle Hinweise und Richtlinien, um diese zum Nutzen des Anwenders zu gestalten [70].

Es gibt zahlreiche, eigens entwickelte Normen zur Gestaltung von Benutzungsschnittstellen, welche Gebrauchstauglichkeit sicherstellen sollen. Anfänglich waren einige dieser Normen auf Bildschirmarbeitsgeräte beschränkt. Später jedoch wurden die meisten Normen auf alle Mensch-Maschine-Systeme ausgeweitet [112, 113, 114, 115]. Bei den Normen wird zwischen denjenigen, die Gebrauchstauglichkeitsanforderungen für Produkte festlegen (sog. Produktnormen) und solchen Normen, welche Entwicklungsprozesse für gebrauchstaugliche Produkte bestimmen (sog. Prozessnormen), unterschieden. Die wichtigsten Normen für gebrauchstaugliche Produkte sind die Produktnormreihe DIN EN 9241 [112] und die Prozessnorm DIN EN ISO 13407 [105].

Die DIN EN 9241 ist die international anerkannte Richtlinie, welche in mehreren Teilen unter dem Namen „Ergonomie der Mensch-System-Interaktion“ die Interaktion zwischen Mensch und Maschine (System) beschreibt, um hierbei gesundheitliche Schäden zu vermeiden und dem Benutzer die Ausführung seiner Aufgaben zu erleichtern. Durch die 2006 erfolgte Umbenennung von „Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten“ auf den derzeitigen Titel wurde der Umfang der Norm von Bürotätigkeiten mit Bildschirmgeräten auf alle Mensch-Maschine-Systeme erweitert.

In den ersten Teilen setzt die Norm Anforderungen an die Hardware verschiedener Schnittstellen und Arbeitsaufgaben fest (Teile 2-9). Aspekte der ergonomischen Gestaltung werden in den Teilen 11-17 behandelt, Teil 110 (ehemals Teil 10) beschreibt die Grundsätze der Dialoggestaltung, welche bei der Interaktion zwischen Benutzer und System erfüllt sein sollten (Aufgabenangemessenheit, Selbstbeschreibungsfähigkeit, Steuerbarkeit, Erwartungskonformität, Fehlertoleranz, Individualisierbarkeit, Lernförderlichkeit).

Die Norm DIN EN ISO 13407<sup>27</sup> gibt eine Anleitung für die „benutzerorientierte Gestaltung interaktiver Systeme“. Diese Gestaltung konzentriert sich darauf, gebrauchstaugliche Systeme zu entwickeln. Die Notwendigkeit für diesen Prozess wird durch die Ziele des Systems festgelegt, z.B. im medizinischen Bereich durch

---

<sup>27</sup> Zu beachten ist, dass die Namensgebung der Normen nicht dem ergonomischen Grundsatz der Unterscheidbarkeit entspricht. Sucht man anstatt der DIN EN ISO 13407 die Norm DIN EN 13407, so findet man die Norm für „Wandhängende Urinale - Funktionsanforderungen und Prüfverfahren“.

die Anforderungen an gebrauchstaugliche Medizinprodukte im Sinne der Gewährleistung von Gesundheit und Sicherheit des Benutzers und des Patienten. Es handelt sich hierbei um multidisziplinäre Aktivitäten, die über menschliche Faktoren und ergonomische Kenntnisse und Techniken ein Wissen beinhalten, das die Effektivität und Effizienz des Systems erhöht, die Arbeitsbedingungen verbessert und möglichen nachteiligen Auswirkungen auf menschliche Gesundheit, Sicherheit und Leistung entgegenwirken kann. Die iterative Vorgehensweise dieses Prozesses wurde bereits in Kapitel 2.1.4 beschrieben. Die DIN EN ISO 13407 ist eine Prozessnorm, die einen Entwicklungsprozess als konform betrachtet, wenn er ihre Empfehlungen erfüllt. Sie kann aber auch als Produktnorm herangezogen werden, denn sie erklärt jedes Produkt, welches nach dem gegebenen Prozess entwickelt wurde, für gebrauchstauglich.

### **2.3.3 Diskussion der Relevanz der Gesetzgebung und Normen für diese Arbeit**

Um der Fragestellung dieser Arbeit zu entsprechen, mussten die Normen bezüglich ihrer Relevanz für eine valide Standardprozedur überprüft werden. Weder die DIN EN IEC 60601-1-6, die DIN EN IEC 62366, noch die DIN EN ISO 14971 liefern genaue Informationen darüber, wie Gebrauchstauglichkeitsprobleme entdeckt und bewertet werden. Vielmehr geben sie nur eine wenig konkrete Richtung vor, an die sich Entwickler für die Gestaltung medizinischer Produkte halten sollen. Diese unkonkrete Beschreibung ist im Sinne der Norm vollkommen gerechtfertigt, da der Gebrauchstauglichkeitsprozess je nach Ziel und Produkt variieren kann. Andererseits muss im Sinne von Validität und Reliabilität der Bewertungsprozess für medizinische Geräte einer immer gleichen Standardprozedur folgen. Die DIN EN ISO 14971 liefert eine sinnvolle Möglichkeit, Risiken nach der Wahrscheinlichkeit und Häufigkeit des Auftretens eines Schadensereignisses und des Schweregrades zu bewerten. Hierbei wird jedoch nicht explizit auf den Charakter von Gebrauchstauglichkeitsproblemen eingegangen.

Die softwareergonomischen Normen sind auf nicht sicherheitskritische Systeme ausgerichtet. Anforderungen an medizinische Geräte gehen weit über die Anforderungen der Geräte aus diesem Bereich hinaus. Dennoch liefern diese Normen, zusammen mit der gängigen Literatur, ein reichhaltiges Methodenrepertoire, um Gebrauchstauglichkeitsprobleme zu entdecken und zu bewerten. Die DIN EN ISO 13407 liefert ein Prozessrahmenwerk, das Anforderungen an den Gestaltungs-



prozess für interaktive Systeme definiert. Die Norm wird in der Entwicklungsphase verwendet und benennt Aspekte, die zu berücksichtigen sind. Allerdings werden keine Methoden und auch keine Anleitung für die Umsetzung der einzelnen Phasen beschrieben. Die Norm bietet die Freiheit, einen Prozess auf organisatorischer Ebene zu definieren. Für eine bewertende Standardprozedur können einige Aspekte der Norm verwendet werden. Sie gilt zwar für in der Entwicklung befindliche Produkte, jedoch werden, bedingt durch die zunehmende Konkretisierung im iterativen Prozess, die entstehenden Produkte fertigen Produkten angeglichen, so dass auch das Rahmenwerk dieser Norm verwendet werden kann.

Etwas konkretere Richtlinien für die Beurteilung der Gebrauchstauglichkeit liefert die DIN EN 9241-11 durch den vorgegebenen Rahmen für die Spezifikation und das Messen der Gebrauchstauglichkeit von Produkten. Sie beschreibt jedoch keine Bewertung der aufgefallenen Probleme, lediglich eine unkonkrete Gewichtung der Gebrauchstauglichkeitskriterien.

Generell sind sowohl die Normen aus dem medizinischen Bereich als auch die Normen aus der Softwareergonomie wenig konkret und geben mehr unspezifische Empfehlungen als Handlungsanweisungen [171]. Dennoch lassen sich durch sie, in Verbindung mit Literatur und Forschung, Ansätze für eine Standardprozedur für die Bewertung der Gebrauchstauglichkeit medizinischer Geräte finden. Für die Entwicklung einer allgemeingültigen Standardprozedur für die Bewertung der Gebrauchstauglichkeit medizinischer Produkte müssen die entsprechenden relevanten Normen verknüpft und mögliche Synergien aufgedeckt und evaluiert werden.

## **2.4 Prüfprozess der Gebrauchstauglichkeit**

Das Thema Gebrauchstauglichkeit rückt immer mehr in das öffentliche Interesse. Canton zeigt beispielsweise, dass in den USA ca. 30% der Home Entertainment-Produkte aufgrund von Bedienproblemen zurückgegeben werden [51]. Bereits seit 1961 werden der Öffentlichkeit Testergebnisse mit Bezug zur Gebrauchstauglichkeit von vergleichenden Warentests durch die Zeitung „DM“ und seit 1966 durch die von der Stiftung Warentest publizierte Zeitschrift „test“ zugänglich gemacht [194]. Durch das öffentliche Interesse und die Informationsverbreitung durch die Medien kommt der Feststellung der Gebrauchstauglichkeit eine erhebliche volks-

wirtschaftliche und werbewirksame Bedeutung zu (welche ausführlich in Kap. 2.2.2 beschrieben wurde).

Es mag zwar sein, dass Gebrauchstauglichkeit teilweise subjektiv, vieldimensional, schwer zu messen und unscharf [6, 223] ist, dennoch hat sich eine Fülle an validen Techniken und Werkzeugen hervorgehoben, um gebrauchstaugliche und sichere Geräte zu entwickeln und zu validieren [97, 223]. Ebenfalls wird die Ermittlung der Gebrauchstauglichkeit durch eine Reihe von Normen [93, 97, 104, 113] standardisiert. Grundsätzlich haben bei der Prüfung der Gebrauchstauglichkeit die in nationalen und internationalen Normen festgelegten Prüfverfahren den Vorrang. Diese lassen jedoch dem Anwender meist einen großen Handlungsspielraum, sodass auf das reichhaltige wissenschaftliche Methoden- und Werkzeugrepertoire zurückgegriffen werden kann, wobei letztlich die Vorgaben der Normen erfüllt werden. Dieses Fehlen fester Richtlinien und Maßskalen darf keinesfalls nachteilig gewertet werden. Bei der Gebrauchstauglichkeitsbetrachtung wird durch die unkonkrete Formulierung eine flexible Anpassung an den immerwährenden und dynamischen technischen Fortschritt vereinfacht. Feste Maßskalen, wie z.B. bei Papierformaten oder Schraubengewinden, würden zwar kurzfristig einen Nutzen stiften, jedoch würden die Normen schnell durch den Stand der Technik überholt werden. Durch die lange Dauer der Umsetzung einer Norm und die hohe Dynamik der industriellen Geräteentwicklung wären Normen, welche die Geräteinteraktion betreffen, ohne ausfüllungsbedürftige Formulierungen oft bereits bei der Veröffentlichung veraltet [6].

Die meisten Methoden zur Gebrauchstauglichkeitsprüfung orientieren sich an den Vorgaben der DIN EN 9241-11. Die Maße für Gebrauchstauglichkeit sind hierbei unter den Oberbegriffen Effektivität, Effizienz und Zufriedenstellung zusammengefasst. Neben den in der Norm definierten Gebrauchstauglichkeitskriterien muss bei der Betrachtung von Gebrauchstauglichkeit, besonders in der Medizin, der Aspekt der Sicherheit zusätzlich berücksichtigt werden. Sicherheit ist die Freiheit von nicht akzeptablen Risiken [97, 108]. In der Medizin wird zwischen Patientensicherheit, Benutzersicherheit und Gerätesicherheit unterschieden. Es kommt vor, dass Sicherheit auch als ein Kriterium der Gebrauchstauglichkeit angesehen wird [401], jedoch wird in dieser Arbeit Sicherheit als ein Ziel der Gebrauchstauglichkeit betrachtet, welches durch die Gebrauchstauglichkeitsmaße Effektivität, Effizienz und Zufriedenheit erreicht wird.

Die Messung von Gebrauchstauglichkeit in der Medizin folgt grundsätzlich denselben Größen wie die Messung der Gebrauchstauglichkeit im Allgemeinen. Hinzu kommt die Beachtung der speziellen Arbeitsumgebung und Situation sowie sicherheitsrelevanter Faktoren im medizinischen Arbeitssystem. Die Auswirkung von Gebrauchstauglichkeitsproblemen auf die Sicherheit muss zusätzlich berücksichtigt werden (Tab. 7). Die für Gebrauchstauglichkeit relevante Norm in der Medizin DIN EN IEC 60601-1-6 [97] greift daher die Anforderungen an die Gebrauchstauglichkeit aus der DIN EN 9241-11 auf und betrachtet diese explizit unter dem Aspekt der Sicherheit. Sie schreibt Herstellern medizinischer Produkte die Validierung der Gebrauchstauglichkeit als letzte Phase des Ergonomieprozesses und deren Dokumentation in einer Ergonomieakte vor.

Effektivität:	Die Genauigkeit und Vollständigkeit, mit der ein Benutzer ein bestimmtes Ziel erreicht. Als Maßstab dienen z.B. die Anzahl der Fehler und die Anzahl der erfolgreich bewältigten Aufgaben. Die Zielerreichung hierbei wird mit einem Maßstab bewertet [113].
Effizienz:	Der im Verhältnis zur Genauigkeit und Vollständigkeit eingesetzte Aufwand, mit dem ein Benutzer ein bestimmtes Ziel erreicht. Als Maßstab dient hierbei z.B. die Zeit, welche benötigt wurde, um eine Aufgabe erfolgreich zu bewältigen. Hierbei muss die relative Zeit bewertet werden, welche der Nutzer tatsächlich zur Aufgabenerfüllung benötigt [113].
Zufriedenheit:	Freiheit von Beeinträchtigungen und positive Einstellungen gegenüber der Nutzung des Produktes. Als Maßstab dient hierbei die Meinung der Nutzer [113].
Sicherheit	Freiheit von unvertretbaren Risiken [108]

**Tab. 7: Gebrauchstauglichkeitskriterien mit Definitionen**

Zur Messung der Gebrauchstauglichkeit muss laut dieser Norm ein Hersteller medizinischer Geräte einen Validierungsplan für Gebrauchstauglichkeit entwickeln, nach welchem er die Konzeption der Schnittstelle zwischen Anwender und Gerät validieren muss. In diesem müssen die verwendete Methode, die wesentlichen Betriebsfunktionen, die Validierungskriterien und die beteiligte Anwendergruppe spezifiziert werden. Die Norm lässt es dem Hersteller jedoch frei, ob die Validierung qualitativ oder quantitativ erfolgt bzw. ob die Validierung im Labor, in einer simulierten oder in der wirklichen Gebrauchsumgebung durchgeführt wird. Der Validierungsplan muss aber Gebrauchsszenarien für den schlechtesten Fall

der Verwendung berücksichtigen. Diese werden von der Spezifikation der Anwendung, vorhersagbaren Gebrauchsfehlern, Ergebnissen der Risikoanalyse und den Gebrauchsszenarien der häufigsten Anwendungssituationen abgeleitet. Die Validierung der Gebrauchstauglichkeit wird von einer Benannten Stelle überprüft [97].

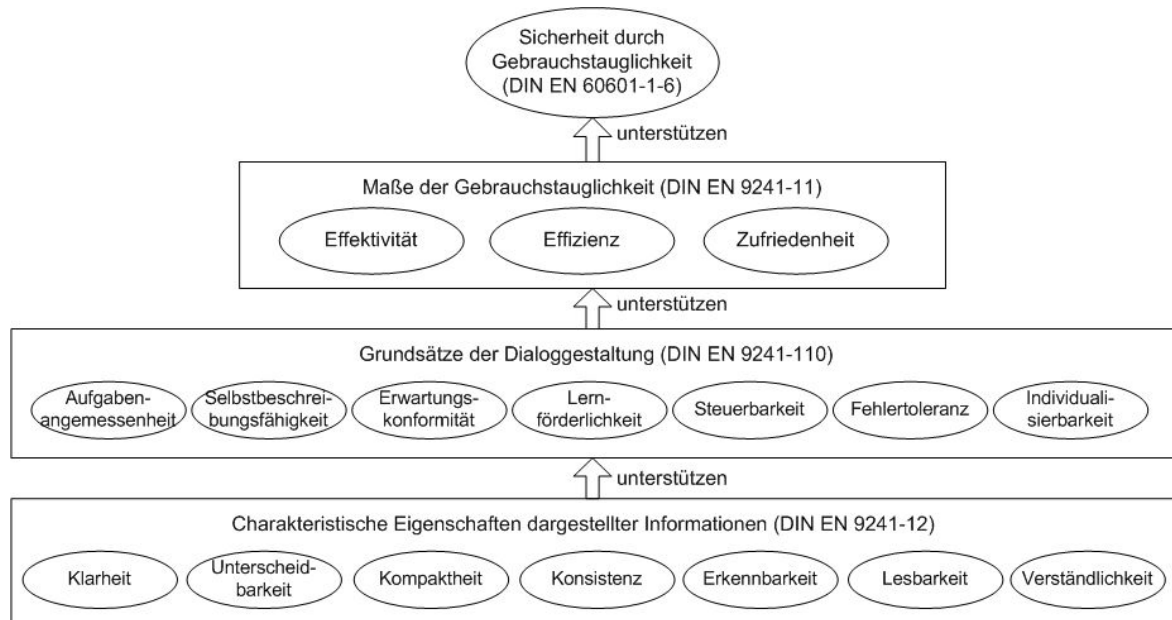
Um Gebrauchstauglichkeit bestimmen zu können, müssen Maßstäbe zur Überprüfung der Gebrauchstauglichkeit aufgestellt und Ziele der Gebrauchstauglichkeit definiert werden. Für die jeweiligen Gebrauchstauglichkeitskriterien gibt es unterschiedliche Maße, die je nach Nutzungskontext und Benutzungsziel des Produktes abweichen können. Um quantitative Aussagen treffen zu können, müssen die Kriterien entsprechend operationalisiert werden [113, 223]. Abb. 24 veranschaulicht beispielhaft die Operationalisierung der Gebrauchstauglichkeit. Für jedes Gebrauchstauglichkeitsmaß können objektive oder subjektive Daten gewonnen werden, die die Gebrauchstauglichkeit eines Produktes bestimmen.



Abb. 24: Operationalisierung der Gebrauchstauglichkeit nach DIN EN 9241-11 [113, 223, 232]

Auch können zur Überprüfung der Gebrauchstauglichkeit die in der DIN EN 9241 Teil 110 definierten Grundsätze für die Dialoggestaltung (Abb. 25) und die im Teil 12 definierten charakteristischen Eigenschaften zur Informationsdarstellung angewendet werden [114]. Diese Grundsätze nehmen auf die im Teil 11 definierten Maße der Gebrauchstauglichkeit entscheidenden Einfluss. Sie stellen Merkmale

eines Systems dar und können mit Zielen der Gebrauchstauglichkeit zusammenhängen, welche als wünschenswerte Produktmerkmale angesehen werden. Wie jedoch diese Gestaltungsgrundsätze bei der Evaluation verwendet werden, hängt von dem jeweiligen Nutzungskontext ab [113]. In einer Gebrauchstauglichkeits-evaluation können sie als Ziele der Gebrauchstauglichkeit überprüft werden.



**Abb. 25: Zusammenhang zwischen der DIN EN IEC 60601-1-6 und speziellen Teilen der DIN EN 9241**

Im folgenden Kapitel werden verschiedene Methoden und Techniken vorgestellt, die den Anspruch haben, Gebrauchstauglichkeit sowohl als komplettes Konstrukt, als auch einzelne Faktoren zu evaluieren oder bei der Evaluation zu unterstützen. Weiterhin wird die Eignung der Methoden für die in dieser Dissertation bearbeitete Fragestellung untersucht.

### 2.4.1 Methoden der Gebrauchstauglichkeitsevaluation

Evaluation ist nach Wottawa „das Sammeln und Kombinieren von Daten mit einem gewichteten Satz von Skalen, mit denen entweder vergleichende oder numerische Beurteilungen erlangt werden sollen“ [422]. Zu einer vergleichbaren Definition kommen Görner & Ilg. Für sie ist Evaluation ein Vorhaben, welches durch ein „systematisches Sammeln, Auswerten und Interpretieren von Daten eine reliable und valide Bewertung der Benutzungsschnittstelle ermöglicht“. In der Ergonomie wird aus den Ergebnissen der Evaluation abgeleitet, ob ein vorab definiertes Designziel erreicht ist, bzw. ob und wo weitere Verbesserungsmöglich-

keiten ausgeschöpft werden können [155]. Für die Gesellschaft für Evaluation e.V. (DeGEval) ist Evaluation die systematische Untersuchung und Bestimmung des Nutzens oder der Güte eines Gegenstandes [24]. Es lässt sich noch eine Fülle an Definitionen in der Literatur finden, mit teilweise kontroversen Inhalt und aus verschiedenen Blickwinkeln erhoben. Auf jeden Fall ist Evaluation charakterisiert durch die Systematik einer Bewertung mit dem Ziel der Erkenntnisgewinnung über das Subjekt der Untersuchung. Sie sollte hierbei die von der Gesellschaft für Evaluation definierten grundlegenden Eigenschaften Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit aufweisen [24]. Grundsätzlich dient sie nie dem Selbstzweck und ist nur dann sinnvoll, wenn auf der Grundlage ihrer Ergebnisse praktische Konsequenzen folgen [421], etwa die Verbesserung oder Veränderung des Interaktionskonzeptes. Des Weiteren muss eine Evaluation, besonders im medizinischen Umfeld, im Sinne der wissenschaftlichen, ethischen und moralischen Verantwortung erfolgen [49, 36, 284, 421]. Darüber hinaus muss sie die Testgütekriterien Validität, Reliabilität und Objektivität erfüllen [199, 36, 168]. Dies kann durch Wiederholbarkeit, Reproduzierbarkeit, Unparteilichkeit und Sachlichkeit erreicht werden. Die gleichen Kriterien gelten auch für die Evaluation der Gebrauchstauglichkeit.

Keine Methode zur Erfassung der Gebrauchstauglichkeit ist für jede Situation optimal geeignet. In der Regel werden unterschiedliche Methoden oder Kombinationen bei der Produktentwicklung und der Gebrauchstauglichkeitsevaluation angewandt. Generell sollten Methoden bevorzugt werden, welche objektive und auditierbare Daten generieren, jedoch können auch subjektive Fakten entscheidende Daten für einen Evaluationsprozess liefern. Dies ist besonders bei der Evaluation der Nutzerzufriedenheit der Fall.

Im Rahmen der Gebrauchstauglichkeitsevaluation sind zwei Evaluationsarten anwendbar. Man unterscheidet in formative (begleitende) Evaluation und summative Evaluation (Kontrolltest) [36, 147, 189]. Die formative Evaluation wird in die Entwicklung integriert. Sie ist durch eine fortlaufende Kontrolle der Maßnahmen bei der Produktentwicklung und deren Auswirkung charakterisiert und stellt den fundamentalen Bestandteil der iterativen Produktentwicklung im Sinne des Usability Engineerings dar (vgl. Kap 2.1.4). Gebrauchstauglichkeitsdefizite können während der Produktentwicklung aufgedeckt und behoben werden. Die summative Evaluation steht hingegen am Ende des Entwicklungsprozesses. Sie dient einer abschließenden Beurteilung des Gerätes. Produktverbesserungen sind in

dieser Phase nicht mehr möglich, jedoch können die Ergebnisse in Nachfolgeprodukten oder späteren Versionen verwendet werden. Bei beiden Evaluationsformen ist es unerheblich, ob es sich hierbei um eine vergleichende Evaluation (deduktiver Ansatz) handelt, mit dem Ziel, unterschiedliche Systeme miteinander zu vergleichen und zu bewerten [42, 43], um eine bewertende Evaluation zur Überprüfung von speziellen Systemeigenschaften, um eine analysierende Evaluation zur Ermittlung von Schwachstellen und entsprechenden Lösungsvorschlägen [173] oder um Kombinationen derselben.

Die für eine Evaluation bereitstehende Methodenvielfalt kann nach verschiedenen Kriterien in Kategorien unterteilt werden: nach Art der Datenerhebung in subjektive und objektive Verfahren, nach den Akteuren in analytische (heuristische) Methoden, die eine leitfadenorientierte Evaluation durch Experten darstellen, und empirische Methoden, welche sich durch Tests mit realen Nutzern auszeichnen, weiterhin nach ihrem Umfang in Teil- und Kompletprüfungen oder nach dem angewandten Beweisprinzip in Verifikation (Nachweis der Fehlerfreiheit) oder Falsifikation (Unterstellung der Fehlerfreiheit bis zum Gegenbeweis) [36, 150, 173].

Inspektionsmethoden	Empirische Methoden
Guideline Review	Videokonfrontation
<b>Heuristische Evaluation</b>	<b>Constructive Interaction</b>
<b>Cognitive Walkthrough</b>	<b>Feldbeobachtung</b>
<b>Pluralistic Walkthrough</b>	Co-discovery Learning
<b>Feature Inspection</b>	Question-Asking-Protocol
Scenario-based Inspection	<b>Nutzertest</b>
Standard Inspection	<b>Lautes Denken</b>
<b>Consistency Inspection</b>	Remote Testing
<b>Checklisten</b>	<b>Retrospective Testing</b>
<b>Fokusgruppe</b>	<b>Coaching Method</b>
	Teaching Method
	Shadowing Method
	Eye Tracking
	<b>Benutzerbefragungen</b>
	<b>Szenario Methode</b>

Tab. 8: Verschiedene Methoden des Usability Engineerings; fettgedruckte Methoden werden im Rahmen dieser Arbeit betrachtet.

Im folgenden Kapitel werden die wichtigsten Methoden der Gebrauchstauglichkeitsevaluation vorgestellt und ihre Relevanz als Evaluationsverfahren im Sinne einer „Standardprozedur für die Gebrauchstauglichkeit von Medizinprodukten“

diskutiert. Hierbei handelt es sich sowohl um einzelne Methodenelemente als auch um Methoden, welche sich verschiedener weiterer Methoden bedienen. Eine Unterscheidung erfolgt hierbei anhand der Evaluatorengruppe in Inspektions- und empirische Methoden (Tab. 8) [183].

#### **2.4.1.1 Gebrauchstauglichkeits-Inspektionsmethoden**

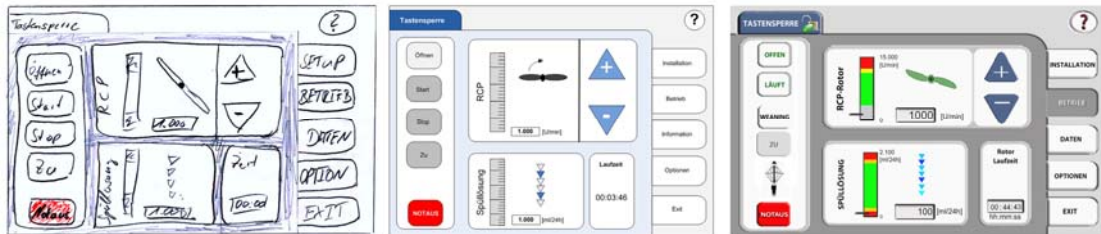
Inspektionsmethoden stellen eine Anzahl von Methoden dar, bei denen Gutachter ein Produkt bezüglich seiner Gebrauchstauglichkeit überprüfen [289]. Generell handelt es sich um Verfahren, die kostenorientiert, auf anerkannten Hypothesen und Regeln (den sog. Heuristiken) basierend, mit relativ wenig Aufwand eine Benutzeroberfläche analysieren und überprüfen, ob geforderte Merkmale vorhanden sind. In manchen Quellen ist auch die Bezeichnung „Heuristische Methoden“ zu finden [212, 223]. Sie unterscheiden sich in Anzahl, Zusammensetzung und Erfahrungsgrad der Experten sowie dem Evaluationsziel [173]. Die gängigsten Methoden sind: Cognitive Walkthrough, Heuristische Evaluation, Pluralistic Walkthrough, Feature Inspektion, Checklisten, Konsistenzprüfung und Fokusgruppen.

##### **Kognitive Überprüfung (engl. Cognitive Walkthrough)**

Die kognitive Überprüfung, besser bekannt unter dem englischen Synonym „Cognitive Walkthrough“, ist eine aufgabenorientierte Inspektionsmethode [159, 65, 173], mit der Experten die Gebrauchstauglichkeit eines Gerätes untersuchen. Die von Lewis et al. [236] veröffentlichte Methode basiert auf der Verwendung von kognitiven Modellen bei einer Benutzerschnittstelle und dem Modell des explorativen Lernens bei der Verwendung eines Produktes. Nutzer versuchen dementsprechend durch schrittweises Probieren, ihr Ziel mit dem Produkt zu erreichen [65, 143, 159, 236, 312, 410]. Die Methode richtet ihren Fokus hierbei weniger auf das Produkt, sondern mehr auf die gedanklichen Prozesse des Anwenders bei der Bedienung [11]. Experten simulieren hierbei einen Nutzer und erkunden so das Gerät und dessen Bedienung [368]. Sie überprüfen die Logik der vorgegebenen Interaktionsschritte und versuchen, Gebrauchstauglichkeitsprobleme aufzudecken. Hierfür erhalten sie detaillierte Angaben über den Nutzungskontext.



Anhand von Mock-Ups<sup>28</sup>, Prototypen (Abb. 26) oder Interaktionskärtchen erlaubt diese Methode bereits im Frühstadium des Designprozesses die Bewertung eines oder mehrerer Designvorschläge [97]. Nach Lewis et al. sollen so bis zu ca. 50% aller Bedienprobleme ermittelt werden können [236].



**Abb. 26: Unterschiedliche Ausprägungen eines Prototypen: Einfache Papierskizze, Drahtmodell und endgültiges Designkonzept**

Ein Cognitive Walkthrough umfasst folgende Schritte: Die Vorbereitungsphase, in der Benutzertypen, Aufgaben und Handlungssequenzen erfasst und beschrieben werden, die Analysephase, in der die jeweiligen Interaktionssequenzen anhand von speziellen Fragestellungen sukzessiv untersucht werden [173], und die Follow-Up-Phase, in der das System entsprechend den Ergebnissen auf die kognitiven Fähigkeiten der Benutzer angepasst wird [173, 410]. Eine Checkliste für die Durchführung dieser Methode ist bei Görner zu finden [155].

Ein Vorteil des Cognitive Walkthrough ist die frühzeitige und kostengünstige Anwendung und der aufgabenzentrierte Ansatz. Hierbei werden neben der Aufdeckung der Nutzerprobleme in den spezifischen Aufgaben auch deren Ursachen erläutert. Die Methode kann in kurzer Zeit relevante Schwächen aufdecken. Ein Nachteil ist die Unklarheit, ob die simulierten Sequenzen angemessen und die Einschätzung des Experten hinsichtlich der Fähigkeiten des Nutzers richtig sind [159, 173]. Anwenderspezifische Defizite werden mit dieser Methode nur selten entdeckt. Die Methode ist im großen Maße abhängig von der Fähigkeit der Experten, sich in die Rolle des Anwenders im Nutzungskontext versetzen zu können.

Fazit: Der Cognitive Walkthrough ist zwar ein geeignetes Verfahren, um in der frühzeitigen Produktentwicklung Gebrauchstauglichkeitsprobleme aufzudecken und die Ergebnisse in die Entwicklung zu integrieren, es ist jedoch für die umfassende Bewertung einer Benutzerschnittstelle nicht hinreichend ausgelegt.

<sup>28</sup> Ein Mock-Up ist ein Vorführmodell, mit dem eine Benutzerschnittstelle für eine Evaluation simuliert werden kann. In der Regel wird er für Prototypen in frühen Stadien eingesetzt, wie z.B. Papierprototypen, Screenshots oder erste Bildschirmkonfigurationen.

### Pluralistic Walkthrough

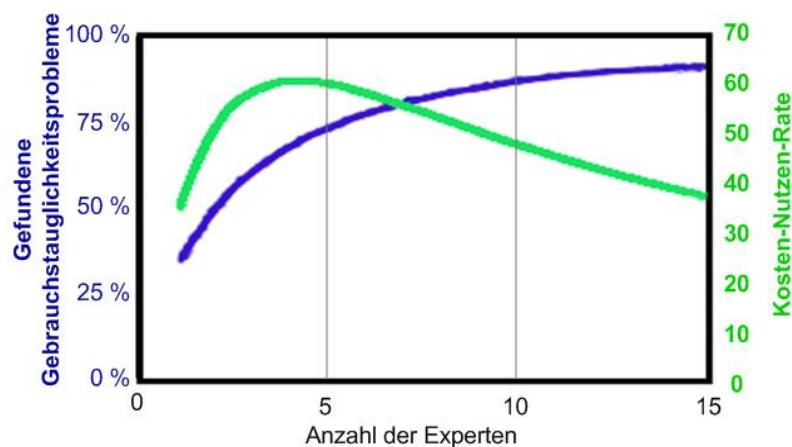
Der Pluralistic Walkthrough ist im Gegensatz zum Cognitive Walkthrough eine Technik, die den Nutzer mit einbezieht. Bei dieser Methode versucht ein interdisziplinäres Team, bestehend aus Entwicklern, zukünftigen Nutzern und Gebrauchstauglichkeitsexperten, Gestaltungsprobleme zu entdecken [26, 65, 118, 338, 379]. Der Pluralistic Walkthrough eignet sich besonders im frühen Stadium der Produktentwicklung für die Erstellung eines benutzerorientierten Designs (vgl. Kap. 2.1.4). Hierbei werden den zukünftigen Nutzern zuerst das System, das Bedienkonzept und die zu erfüllenden Aufgaben in Form von Szenarien (► Kap. 4.1.5) präsentiert. Sie versuchen danach, die gestellten Aufgaben, dem Entwicklungsstand entsprechend, mit Papierprototypen, Wandzeichnungen der jeweiligen Screens, Mock-Ups oder weiter fortgeschrittenen Prototypen zu erfüllen. Die Entwickler fungieren hierbei als eine Art „lebendige Hilfe“ und der Gebrauchstauglichkeitsexperte als Moderator. Nachdem die Nutzer die Aufgaben bewältigt haben, stellen sie selbst ihre Lösung vor, worauf ihnen die „richtige“ Lösung präsentiert wird. Danach werden Benutzerprobleme und Lösungen diskutiert.

Der Vorteil dieser Methode liegt darin, dass Entwickler eine direkte Rückkopplung durch die Benutzer erhalten, einen Einblick in Benutzerprobleme bekommen, und dass Gebrauchstauglichkeitsprobleme vor dem Hintergrund verschiedener Erfahrungen, Fähigkeiten und Sichtweisen identifiziert werden können. Allerdings stellt diese Methode hohe Ansprüche an den Moderator, da er erreichen muss, dass sich die Nutzer, konfrontiert mit den Entwicklern, trauen, Kritik an der Benutzerschnittstelle vorzubringen. Andererseits besteht das Problem, dass die Entwickler die Kritik neutral annehmen und nicht in ihrem Sinne abwehren oder verharmlosen [338, 350]. Nachteilig ist einerseits der hohe Aufwand, da pro Benutzungspfad ein Walkthrough durchgeführt werden muss, andererseits die Einschränkung, dass die Anwender das Gerät nicht wirklich verwenden, also keine typischen Aufgaben erleben können.

Fazit: Der Pluralistic Walkthrough ist eine sinnvolle Methode im frühen Stadium einer Entwicklung, jedoch für eine alleinige Bewertung zu aufwändig und unzureichend.

## Heuristische Evaluation

Die Heuristische Evaluation geht von der Gültigkeit bekannter Gestaltungsregeln [65, 97, 118, 223, 349] und Standards (den sog. Heuristiken) aus und überprüft deren Einhaltung bei einem Produkt mit Hilfe von Expertengutachtern. Die Experten simulieren die Sicht der Benutzer, mit dem Ziel, Probleme aufzudecken und gute Gestaltungslösungen der Benutzeroberfläche zu finden. Untersuchungen von Nielsen haben gezeigt, dass bereits eine geringe Anzahl von Experten für die Evaluation einer Benutzerschnittstelle ausreicht [223, 282, 284, 290]. Sein Ergebnis ist, dass bereits fünf Experten ca. 75 % der Gebrauchstauglichkeitsprobleme aufdecken. Bei entsprechender Erhöhung der Evaluatorenanzahl steigt die Zahl der gefundenen Probleme nur unwesentlich, und dementsprechend ist das Aufwand-Nutzen-Verhältnis nicht mehr angemessen (Abb. 27). Für sicherheitskritische Bereiche oder speziell für den medizinischen Bereich gibt es jedoch bisher keine derartigen Untersuchungen in der Literatur.



**Abb. 27: Gefundene Probleme und Kosten-Nutzen-Verhältnis in Abhängigkeit von der Expertenzahl bei einer Heuristischen Evaluation [284, 289]**

Als Evaluationsgrundlage für eine heuristische Evaluation können neben den entsprechenden Normen bekannte Checklisten eingesetzt werden, anhand derer die Experten in oft mehreren Durchläufen (ca. 2-3) die Oberfläche inspizieren. So haben beispielsweise Nielsen & Molich aufgrund ihrer Erkenntnis, dass Gebrauchstauglichkeitsprobleme bestimmten Kategorien zuzuordnen sind, eine Checkliste entwickelt [271], deren Überarbeitung in zehn allgemeine Heuristiken resultierte [282, 284]:

- **Sichtbarkeit des Systemstatus:** Ein System sollte den Benutzer jederzeit angemessen darüber informieren, was passiert.

- **Übereinstimmung zwischen System und Wirklichkeit:** Ein System sollte die Sprache des Benutzers sprechen, in einer Form, mit der der Benutzer vertraut ist, und nicht mit systemorientierten Begriffen. Informationen sollten in natürlicher und logischer Reihenfolge dargestellt werden.
- **Nutzerkontrolle und -freiheit:** Benutzer wählen Systemfunktionen oft versehentlich und benötigen somit einen klar gekennzeichneten »Notausgang«, um den ungewünschten Zustand ohne langen Dialog verlassen zu können. Die Funktionen »Rückgängig« und »Wiederholen« sollten unterstützt werden.
- **Konsistenz und Standards:** Benutzer sollten sich nicht fragen müssen, ob verschiedene Wörter, Situationen oder Aktionen dasselbe meinen. Plattformkonventionen sollten befolgt werden.
- **Fehlervorbeugung:** Ein sorgfältiges Design, welches das Auftreten von Problemen verhindert, ist noch besser als gute Fehlermeldungen.
- **Erkennen anstatt Erinnern:** Objekte, Aktionen und Optionen sollten sichtbar sein. Der Benutzer sollte sich Informationen nicht von einem Abschnitt des Dialogs bis zu einem anderen merken müssen. Instruktionen für den Systemgebrauch sollten leicht auffindbar sein.
- **Flexibilität und Effizienz:** Akzeleratoren (»Programmzeitverkürzer«) können (von Laien unbemerkt) die Interaktion für Experten so beschleunigen, dass ein System sowohl von Anfängern als auch von erfahrenen Benutzern bedient werden kann. Benutzern sollte ermöglicht werden, häufige Aktionen auf sie zuzuschneiden.
- **Ästhetisches und minimalistisches Design:** Dialoge sollten keine Informationen beinhalten, die irrelevant sind oder selten benötigt werden. Jede zusätzliche Informationseinheit in einem Dialog konkurriert mit den relevanten Informationen.
- **Hilfe beim Erkennen, Diagnostizieren und Bewältigen von Fehlern:** Fehlermeldungen sollten in einfacher Sprache formuliert sein, das Problem exakt beschreiben und eine konstruktive Lösung vorschlagen.
- **Hilfe und Dokumentation:** Obwohl es besser ist, wenn ein System ohne Dokumentation verwendet werden kann, mag es notwendig sein, Hilfe und Dokumentation bereitzustellen. Diese Informationen sollten leicht zu durchsuchen und auf die Aufgabe des Benutzers fokussiert sein; zudem sollten sie konkrete Schritte, die vorgenommen werden müssen, aufzählen und dabei nicht zu umfangreich sein.

Weitere Heuristiken sind bei Tognazzini [392], Shneidermann [368], Pierotti [311], Purho [319] oder speziell für ältere Benutzer (50+), bei der amerikanischen „Association for the Advancement of Retired Persons“ (AARP) zu finden [62].

Der Vorteil dieser informellen Methoden des Usability Engineerings ist, dass sie eine im Vergleich vor allem zu den empirischen Methoden kostengünstige Alternative [143] darstellen, die schnell zu Ergebnissen führt, viele Probleme aufdeckt, keine aufwändige Test-Infrastruktur fordert und frühzeitig im Entwicklungsprozess eingesetzt werden kann. Entwickler und Experten erhalten eine methodische Vorgabe, um ein Gerät zu evaluieren. Nachteilig bei der Heuristischen Evaluation ist, dass die Gutachter keine wirklichen Nutzer sind. Sie können zwar versuchen, sich in die Rolle des Nutzers zu versetzen, werden diesen Status jedoch niemals erreichen. Daher weichen meist die durch die „echten“ Nutzer und die durch Experten entdeckten Probleme voneinander ab [42, 43, 72, 73, 74, 120, 205]. Weiterhin sind Experten häufig durch bereits bekannte Geräte geprägt. Auch werden anwenderspezifische Probleme nur schwer entdeckt [187] und der Nutzungskontext wird weitgehend vernachlässigt [11, 282]. Laut Krömker kann eine Heuristische Evaluation, besonders bei Applikationen und Benutzern, deren Aufgaben einen hohen Professionalisierungsgrad haben, für Gebrauchstauglichkeitsexperten alleine zu schwierig werden [223], weshalb es in diesem Fall wichtig ist, auch Fachexperten hinzuzuziehen. Um Beeinflussungen zu vermeiden und eine Unabhängigkeit zu gewährleisten, ist es vor allem notwendig, dass die Gutachter erst alleine das Gerät überprüfen und dann im Anschluss die Ergebnisse diskutieren.

Eine besondere Form der Heuristischen Evaluation ist die von Muller et al. beschriebene Participatory Heuristic Evaluation, die auch Nutzer in den Kreis der Gutachter einbezieht [279]. Diese Erweiterung löst jedoch nicht das Problem einer unzureichenden Nutzermeinung, da sich diese Nutzer nicht in einer Arbeitsumgebung befinden und sich daher auch einiger Probleme nicht bewusst werden können. Außerdem wären Nutzer in der ungewohnten Evaluationssituation eventuell voreingenommen oder eingeschüchtert.

Fazit: Die Heuristische Evaluation ist ein geeignetes, kostenorientiertes Verfahren, um Gebrauchstauglichkeitsprobleme zu evaluieren. Aufgrund der Problematik, dass Experten nicht Nutzer sind, kann sie nicht als einziges Evaluationsverfahren für die Bewertung einer Benutzerschnittstelle eingesetzt werden, wohl aber ergänzend, um über das Gerät einen ersten Eindruck zu bekommen, auf dem die eigentliche Bewertung aufbaut.

### **Funktionsanalyse (Feature Inspection)**

Bei der Funktionsanalyse werden durch Experten spezifische Funktionalitäten (engl. features) auf mögliche Probleme hin untersucht. Hierbei werden die Aufgaben, die ein Nutzer mit dem Gerät typischerweise durchführen wird, im Hinblick auf zu lange, umständliche, unnatürliche oder unnötige Aktionsfolgen hin analysiert [284]. Durch diese kostengünstige Methode können zwar einige Gebrauchstauglichkeitsprobleme gefunden werden, jedoch wird auch hier die Rolle des Nutzers und der Arbeitsumgebung vernachlässigt.

Fazit: Die Funktionsanalyse bietet eine Möglichkeit, einige Gebrauchstauglichkeitsprobleme aufzudecken, sie ist jedoch für die Bewertung eines Systems nicht ausreichend.

### **Checkliste**

Eine Checkliste ist ein Prüffragenkatalog zur Aufnahme eines bestimmten Zustandes. Sie liegt einer Operationalisierung („Messbarmachung“) von speziellen Normen zugrunde [6]. Checklisten werden meistens von Arbeitsmedizinern und Ergonomen verwendet, um einen Arbeitsplatz erst neutral zu erfassen und später mit erarbeiteten Standards und Normen abzugleichen [174, 189, 221]. Eine Checkliste für interaktive Anwendungen ist bei Holzinger [189] und eine speziell für die Bewertung des Arbeitsplatzes Chirurgie bei Koneczny zu finden [221]. Der Vorteil von Checklisten ist, dass sie einfach und kostengünstig durchzuführen sind, jedoch berücksichtigen sie nicht ausreichend die Gebrauchstauglichkeit.

Fazit: Eine Checkliste ist geeignet für die reine Aufnahme eines Zustandes. Sie lässt jedoch eine Einbeziehung der Nutzer vermissen und ist daher für eine Gebrauchstauglichkeitsbewertung ungeeignet.

### **Fokusgruppe**

Eine Fokusgruppe zählt zu den qualitativen Methoden der Datenerhebung [173]. Sie ist eine Gruppendiskussion [124, 278], in der eine Gruppe von Personen in strukturierter oder moderierter Weise über ein bestimmtes Thema diskutiert [36]. Ziel ist dabei, mit einer kleinen Anzahl von Teilnehmern ein vorgestelltes Produktkonzept oder ein bestehendes Produkt zu bewerten [36, 41]. Dabei werden die Teilnehmer mit den verschiedenen Sichtweisen, Ideen und Wahrnehmungen der

anderen Gruppenteilnehmer konfrontiert und tauschen sich innerhalb einer zeitlich begrenzten Diskussion [130] untereinander aus.

Der Vorteil einer Fokusgruppe ist der, dass in kurzer Zeit und mit vertretbarem Aufwand Kundenmeinungen und Anregungen gesammelt werden können, um auf diese Weise eine bessere Grundlage für die Entwicklungsentscheidung zu gewinnen [173]. Bei Beteiligung der Nutzer können Gesichtspunkte betrachtet werden, die den Experten verborgen bleiben könnten. Ein Nachteil ist, dass durch gruppenspezifische Effekte Einzelmeinungen dominieren können. Auch bleiben unbewusste Probleme der Nutzer verborgen.

Fazit: Eine Fokusgruppe ist nur möglich bei bereits auf dem Markt befindlichen Produkten oder weit in der Entwicklung stehenden Prototypen. Ansonsten müsste zuerst eine Produkteinweisung erfolgen.

### **Konsistenzprüfung (engl. Consistency Inspection)**

Die Konsistenzprüfung wird meistens bei komplexen Geräten, Geräteverbänden oder ganzen Systemen verwendet, um Konsistenz und Einheitlichkeit innerhalb des Produktes und des Systems zu überprüfen. Es wird geprüft, ob Funktionen in allen Kontexten gleich ausgeführt werden können [173], und ob Terminologie, Farben, Layout etc. sowohl im Interface als auch in Schulungsunterlagen gleichermaßen konsistent sind [371].

Fazit: Die Konsistenzprüfung sollte bei Systemen in Systemverbänden angewendet werden.

### **2.4.1.2 Empirische Methoden**

Empirische Methoden sind experimentell und werden mit echten Benutzern oder potentiellen Benutzern durchgeführt. Im Gegensatz zu den Inspektionsmethoden sind sie in der Regel aufwändiger. Dennoch werden sie in den meisten Fällen dem Aufwand gerecht, da genauere Ergebnisse zu erwarten sind. Durch eine Einbeziehung der Nutzer kann aufgezeigt werden, was sie denken, erwarten und was in ihnen vorgeht. Empirische Methoden erforschen meist die Gesamtheit der Aufgaben und stellen Zusammenhänge zwischen den Anwendern, Problemen und dem Nutzungskontext dar. Typische empirische Methoden der Gebrauchstauglichkeitsbetrachtung sind: Feldbeobachtung, Nutzertest und Lautes Denken.

### **Feldbeobachtung**

Bei Feldbeobachtungen [36, 105, 225] (engl. Field observation), auch bekannt als „Remote-Testing“, „Live Observation“ oder „Reality Tracking“, integriert der Usability Experte sich selbst in den Nutzungskontext, nimmt am sozialen Umfeld teil und hält dabei nach besonderen Ereignissen, Verhaltensmustern und Problemen Ausschau. Ebenso bieten sich die Möglichkeiten, den Nutzer selbst an seinem Arbeitsplatz mit seinen Erwartungen und Bedürfnissen zu evaluieren. Aus diesem Grund werden Feldbeobachtungen bei der Produktentwicklung nicht nur in der Evaluationsphase durchgeführt, um zu überprüfen, ob das Produkt die Anforderungen der Benutzer, der Arbeitsaufgaben und der Umgebung erfüllt [105], sondern auch in der Konzeptphase [97]. Die Menge der gesammelten Daten und die Reproduzierbarkeit sind hierbei meist von geringerer Bedeutung. Wichtiger sind die erfahrenen Einblicke in den Nutzungskontext [215]. Der Beobachter muss sich jedoch so unauffällig wie möglich verhalten, da jegliche Art von Störung zu einer Verfremdung der Nutzungssituation und möglicherweise zur Verfälschung der Ergebnisse führen kann. Die Feldbeobachtung ist in den meisten Fällen die am einfachsten anwendbare Evaluationsmethode [190]. Die größte Stärke dieser Methode sind die Einblicke in die reale Nutzungsumgebung. Nachteilig ist, dass eine spezielle Auswahl der zu erledigenden Arbeitsschritte nicht möglich ist, und ein Beobachter in den meisten Fällen als ein Fremdkörper im Nutzungskontext angesehen wird.

Ein besonderer Fall ist die teilnehmende Beobachtung. Bortz & Döring sprechen von einer teilnehmenden Beobachtung (engl. Live Observation), „wenn der Beobachter selbst Teil des Geschehens ist, wenn er also seine Beobachtung nicht als Außenstehender macht“ [36]. Hierbei nimmt der Gebrauchstauglichkeitsexperte keine rein beobachtende Rolle ein, sondern nimmt selbst aktiv an der Nutzungssituation teil, in der das Produkt verwendet wird, und sammelt im Hintergrund Daten. Somit wird eine persönliche Beziehung zu dem Beobachteten aufgebaut und man erhält (wenn man als Teil des Geschehens akzeptiert wird) Einblicke, die dem Außenstehenden verborgen bleiben [46]. Eine Schwierigkeit dieser Beobachtungsmethode liegt in der Gradwanderung des Usability Experten, als teilnehmender Beobachter einerseits integriert zu werden und andererseits den „natürlichen“ Ablauf des Geschehens durch eigene Initiative und Aktivitäten nicht zu stören. Weiterhin ist eine genaue Protokollierung des Erlebten für den Beobachter schwierig. Hierbei können Audio- und Videoaufzeichnungen hilfreich sein.



Fazit: Die Feldbeobachtung bietet eine gute Möglichkeit zur Ermittlung des Nutzungskontextes, ist jedoch für die Bewertung eines Produktes erst nach Freigabe (Release) desselben möglich, da nicht zertifizierte Geräte in der Medizin nicht verwendet werden dürfen (► Kap. 2.3). Weiterhin können keine speziellen Aufgaben gestellt werden, und der Beobachter ist abhängig von der vorliegenden Arbeitssituation. Im Spezialfall der teilnehmenden Beobachtung müsste ein Beobachter ausgewählt werden, der für den medizinischen Kontext eine entsprechende Ausbildung hat.

### **Nutzertest**

Unter einem Nutzertest (auch Gebrauchstauglichkeitstest, Usability Test, Laborexperiment oder Benutzertest genannt) wird die empirische Evaluation der Gebrauchstauglichkeit eines Produktes verstanden. Bei dieser Methode werden unter Mitwirkung von Repräsentanten der Nutzergruppe, in einer möglichst repräsentativen Umgebung, mit repräsentativen Aufgaben die Stärken und Schwächen eines Produktes systematisch evaluiert [5, 97, 125, 212, 240, 284]. Hierbei wird versucht, den Nutzungskontext in einem Labor möglichst realistisch darzustellen. Es ist aber auch möglich, diese Tests in der realen Nutzungssituation durchzuführen.

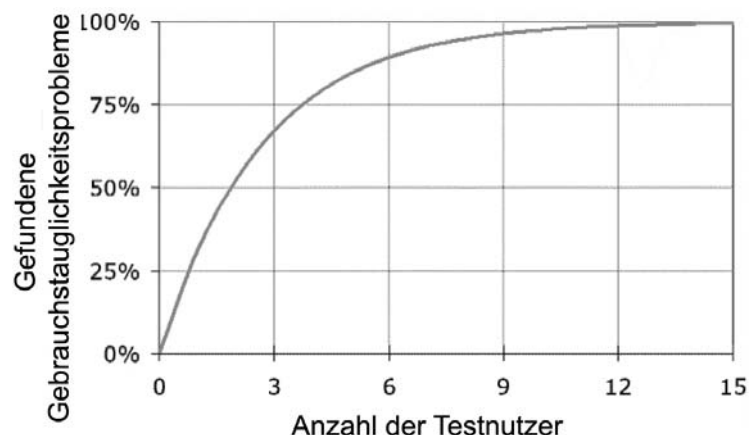
Usability Labore sind grundsätzlich in zwei Bereiche unterteilt. Im Testbereich wird die eigentliche Nutzungssituation simuliert. Aufzeichnungsgeräte wie Videokameras und Mikrofone sollten hier möglichst unauffällig positioniert sein, damit die Versuchsperson ungestört mit dem zu untersuchenden Gerät arbeiten kann. Im Beobachtungsbereich befinden sich Aufzeichnungsgeräte und eine Gegensprechanlage. Von hier aus, meist hinter einem halbdurchlässigen Spiegel oder einem Fenster, können Versuchsleiter, Protokollant und ggf. weitere Interessenten den Versuchsteilnehmer beobachten (Abb. 28). Die Versuchsperson wird auf diese Weise am wenigsten gestört, und es können trotzdem Hilfestellungen und Anweisungen gegeben werden [175]. Für den medizinischen Bereich eignet sich für Nutzertests der europaweit einzige Experimental-OP in Tübingen.



**Abb. 28: Typischer Nutzertest im Experimental-OP in Tübingen**

Während eines Nutzertests werden die Probanden dazu angehalten, typische Arbeitsaufgaben, welche eine Realanwendung reflektieren, mit dem Gerät durchzuführen. Aus den Beobachtungen und Aussagen bei der Aufgabenerfüllung können Rückschlüsse auf die Gebrauchstauglichkeit des Gerätes gezogen werden. Hierzu werden die Maßstäbe Effektivität, Effizienz und Zufriedenheit so operationalisiert, dass qualitative Aussagen möglich sind (vgl. Kap. 2.4).

Nutzertests gehören zu den weit verbreitetsten und zuverlässigsten Methoden, um die Gebrauchstauglichkeit eines Produktes zu ermitteln [6, 11, 124, 125, 212, 284, 314, 336, 349, 352]. Bereits wenige Probanden finden relativ viele Probleme. Laut Nielsen reichen bereits sechs Testnutzer aus, um 80% der Gebrauchstauglichkeitsprobleme zu finden (Abb. 29) [288]. Diese Aussage ist von einigen Experten umstritten und muss für den medizinischen Kontext noch überprüft werden (► Kap. 4.1.4).



**Abb. 29: Prozentzahlen der gefundenen Gebrauchstauglichkeitsprobleme in Abhängigkeit von der Testnutzeranzahl bei einem Nutzertest [288]**

Ein Vorteil dieser Methode ist, dass reale Nutzer typische Aufgaben erfüllen, und dass so die gesamte Erfahrungswelt der Benutzer in die Evaluation einfließen kann. Es können präzise dem Evaluationsziel entsprechende Aufgaben gestellt werden [223]. Es können Probleme aufgedeckt werden, die der Anwender entweder gar nicht bemerkt, wegen einer Routine nicht mehr bemerkt oder zwar schon bemerkt, jedoch nicht als ein Problem identifiziert [253]. Schwachstellen können eindeutig nachgewiesen [336] und Gütekriterien wie Objektivität, Reliabilität und Validität können weitgehend eingehalten werden [336]. Negativ ist jedoch, dass unter Laborbedingungen niemals der vollständig reale Nutzungskontext simuliert werden kann, und somit auch diese Ergebnisse nicht zu 100% auf den realen Nutzungskontext übertragbar sind. Dafür können in einem Labor Störfälle und Stresssituationen simuliert werden, die im klinischen Routinebetrieb aus ethischen Gründen nicht simuliert werden dürfen.

Weiterhin ist es für viele Versuchsleiter schwierig, in der rein passiven Rolle zu bleiben. Darüber hinaus stellt die Schaffung immer gleicher Anfangs- und Testbedingungen ein Problem dar.

Fazit: Der Nutzertest stellt eine Methode dar, die viele Probleme aufdeckt. Mit dieser Methode kann der Nutzungskontext im Labor nachgestellt werden, ohne den Patienten zu gefährden.

### **Lautes Denken**

Die Methode des lauten Denkens (engl. Thinking Aloud) ist eine Methode zur Erfassung bewusster handlungsbegleitender Kognitionen und Emotionen während eines Nutzertests [54, 192, 208, 284, 287, 235, 314], von denen aus auf Problempunkte und bewährte Lösungen der Applikation geschlossen werden kann [46, 147]. Die Anwendung dieser Methode sieht vor, dass der Versuchsteilnehmer bei der Arbeit mit dem zu untersuchenden Gerät seine handlungsbegleitenden Gedanken, Erwartungen und Gefühle verbalisiert. Hierzu ist vor der Untersuchung eine genaue Instruktion notwendig. Diese Äußerungen werden zur späteren Analyse aufgezeichnet. Da der Testnutzer sich im Verlauf des Tests nicht immer an die vorher gegebenen Instruktionen erinnern kann, ist es oft nötig, ihn durch kurze Hinweise wie: „Was denken Sie gerade?“ oder: „Was erwarten Sie?“ zum lauten Denken zu animieren. Bei dieser Methode wird ersichtlich, an welchen Stellen der Nutzer die Applikation falsch interpretiert, und warum diese Fehlinter-

pretation aufgetreten ist. Weiterhin können auf diese Weise Meinungen des Nutzers während des Tests erkannt und gespeichert werden. Als Schwierigkeit wird von Bortz & Döring das „Problem der sozialen Erwünschtheit“ beschrieben [36]. Der Testnutzer glaubt fälschlicherweise, bestimmte Erwartungen erfüllen zu müssen. Deshalb ist es sehr wichtig, der Testperson vor dem Test verständlich zu machen, dass nicht sie getestet wird, sondern das zu verwendende Gerät. Weiterhin sollte man von der Illusion Abstand nehmen, dass man mit dieser Methode das Innere der Testperson nach außen kehren kann. Testpersonen haben erhebliche Schwierigkeiten, ihre Gedanken fortlaufend zu verbalisieren [123, 173]. Die nach Hackos zu erwartenden Einblicke in Schlussfolgerungen, Intuitionen und mentale Modelle [160] werden nur in den seltensten Fällen vollständig erfüllt. Problematisch bei der Methode des lauten Denkens ist für die Probanden, dass sie sich bei einer Verknüpfung mit objektiven Methoden, wie z.B. Performance-Messungen, während der Aufgabenerfüllung auch noch auf das laute Denken konzentrieren müssen. Eine Möglichkeit, dies zu vermeiden, ist die von Hackos vorgeschlagene Methode „Thinking right after the task“ [160], wenn sie auch nicht direkt die Nutzergedanken wiedergibt und auch nicht aufgrund der zeitlichen Spanne und der Beschränktheit des menschlichen Gedächtnisses jeden gehalten Gedanken. Zumindest lenkt sie gerade bei kognitiv anspruchsvollen Aufgaben den Benutzer nicht bei der Aufgabenerfüllung ab. Trotz gewisser Mängel wird diese Methode von den meisten Experten als eine der stärksten Methoden des Usability Engineerings angesehen [160, 190, 284].

Fazit: Das „Laute Denken“ ist eine gute Methode, um einen Einblick in Gedanken und Gefühle sowie mentale Modelle des Benutzers während der Aufgabenerfüllung zu bekommen.

### **Retrospective Testing**

Das Retrospective Testing kann angewandt werden, wenn auf die Methode des lauten Denkens verzichtet wird. Gebrauchstauglichkeitsexperten und Versuchsteilnehmer analysieren nach einem durchgeführten Test zusammen die aufgezeichneten Videosequenzen [287]. Um den bei dieser Methode erheblichen zeitlichen Mehraufwand zu verringern, können zuvor die wichtigen Stellen der Aufnahme („critical incidents“) herausgesucht werden. Die Versuchsperson kann

an diesen wichtigen Stellen ihre Verhaltensweisen erklären, sodass der Experte ähnliche Eindrücke erhält wie beim „lauten Denken“.

Fazit: Das Retrospective Testing ist eine gute Methode, um einen Einblick in Gedanken und Gefühle sowie mentale Modelle des Benutzers nach der Aufgabenerfüllung zu bekommen. Der erhebliche zeitliche Mehraufwand muss aber gerechtfertigt sein.

### **Coaching Methode**

Im Gegensatz zu den meisten empirischen Gebrauchstauglichkeitsevaluationsmethoden ist bei der Coaching Methode eine Interaktion zwischen dem Benutzer und dem Gebrauchstauglichkeitsexperten (dem „Coach“) explizit gewünscht. Bei der Coaching Methode [249, 284, 287] führen Beobachter und Nutzer den Test gemeinsam durch. Der Nutzer wird hierbei ausdrücklich dazu aufgefordert, Fragen an den Coach zu stellen. Ziel ist es, auf diese Weise Auskünfte über den Informationsbedarf der Benutzer zu generieren und diese Informationen in die Gestaltung von Benutzerinformationen, Gebrauchsanleitungen und Benutzerschulungen einfließen zu lassen. Nielsen sieht in dieser Methode darüber hinaus noch den Vorteil, dass bei Produkten, bei denen Testpopulationen rar und schwer zu erreichen sind, Evaluationen in Trainings- und Produktschulungen integriert werden können [284], und somit eine synergetische Situation geschaffen wird. Voraussetzung hierfür ist, dass der Trainer speziell auch als Beobachter geschult sein muss, und der Benutzer noch nicht viel Erfahrung in der Benutzung des Produktes hat. Eine spezielle Form der Coaching Methode ist, dass der Experte und ein erfahrener Benutzer die Rollen tauschen, der Benutzer also die Rolle des Coaches übernimmt. Der Vorteil liegt darin, dass der Beobachter spezielle Einblicke in eine Nutzer-Nutzer-Kommunikation erhält. Diese Methode eignet sich besonders bei Testpersonen, die zurückhaltend mit ihrer Kritik sind, welches in einigen Kulturkreisen (z.B. Japan) mentalitätsbedingt der Fall ist.

Fazit: Die Coaching Methode bietet gute Informationen für die Gestaltung des Begleitmaterials und für Schulungen, aber zu wenig Informationen über Nutzerprobleme. Sie ist daher als Bewertungsmethode ungeeignet.

## **Benutzerbefragung**

Benutzerbefragungen dienen der Evaluation der subjektiven Zufriedenstellung der Benutzer sowie der Feststellung von Ursachen für mangelnde Zufriedenstellung. Sie können nicht nur zum Aufdecken von Benutzungsproblemen im Feld, sondern auch als Methoden für eine Gebrauchstauglichkeitsmessung angewendet werden. In der Benutzerbefragung erhält der Ergonomieexperte Informationen über die Anwender, deren Bedürfnisse, die Nutzungsumgebung, häufig vorkommende Fehler, erlebte Nutzungsprobleme und typische Arbeitsaufgaben. Auch können in der Praxis auftretende Probleme evaluiert und Zufriedenheit gemessen werden. In der Benutzerbefragung werden mit speziellen Fragen systematisch Informationen über den Benutzer gewonnen. Als klassisches Instrument der empirischen Sozialforschung unterscheidet man bei der Befragung nach dem Kommunikationsmedium (online, telefonisch, schriftlich oder persönlich), dem Grad der Standardisierung und nach dem Aufbau der Befragung. Als oberstes Differenzierungsmerkmal kann die Art der Befragung angesehen werden, weshalb man grundsätzlich Interviews und Fragebögen unterscheidet [36].

### **a. Benutzerinterview**

Benutzerinterviews entsprechen einer mündlichen Befragung von Benutzern, in welcher deren subjektive Sichtweise ermittelt wird. Es gibt eine Vielzahl von Interviewtechniken. Obwohl Unterscheidungen nach Art des Kontaktes, Anzahl der befragten Personen, Anzahl der Interviewer und der Funktion [36] möglich sind, unterscheiden sie sich grundsätzlich in dem Grad der Standardisierung. Interviews können vollkommen offen (unstrukturiert) bis hin zu komplett standardisiert (strukturiert) ablaufen. In der Gebrauchstauglichkeitsevaluation können Interviews nicht nur als eigenständige Evaluationsmethode (z.B. in der Anforderungsanalyse), sondern auch als ergänzendes Tool zu einer anderen Methode für die Evaluation der subjektiven Benutzermeinung eingesetzt werden. Nielsen ist der Auffassung, dass Interviews im besonderen Maße dazu geeignet sind, im Anschluss an die Nutzung die subjektive Zufriedenheit der Probanden sowie ihre persönliche Einschätzung zu den einzelnen Aspekten der Gebrauchstauglichkeit zu erheben [284]. Ein Vorteil von Interviews ist die hohe Flexibilität [284, 352]. Im Verlauf solcher Interviews ist es u.a. möglich, gezielt nachzufragen, wenn bestimmte Aktionen oder Aussagen der Probanden unverständlich geblieben sind [117]. Die von Bortz & Döring erwähnten möglichen methodischen Mängel und somit ent-

stehenden Einbußen in den testtheoretischen Gütekriterien [36] können durch einen erhöhten Grad an Standardisierung verringert werden. Es besteht jedoch weiterhin die Gefahr einer Anfälligkeit der Resultate durch Besonderheiten des Interviewers und der Befragungssituation.

Fazit: Das Benutzerinterview stellt eine gute Methode zur Evaluation der subjektiven Benutzermeinung dar. Es ist besonders geeignet in den Phasen der Anforderungsanalyse und als ergänzendes Tool für weitere Evaluationsmethoden.

### **b. Fragebogen**

Benutzerfragebögen entsprechen einer schriftlichen Befragung von Benutzern. Diese Variante erfordert eine hohe Strukturierbarkeit und verzichtet im Gegensatz zu einem Interview auf Eingriffe des Interviewenden [36]. Beim Usability Engineering kommen Fragebögen als Erhebungsmethoden und als unterstützende Tools anderer Erhebungsmethoden zum Einsatz. Hierbei werden hauptsächlich Einstellungsfragebögen [36] zur Ermittlung der subjektiven Nutzerzufriedenheit, Fragebögen zu Gebrauchstauglichkeitsproblemen und Nutzeranforderungen sowie Fragebögen zur Ermittlung der Benutzerdaten und der Daten des Nutzungskontextes verwendet.

Wottawa fordert aus Gründen der Exaktheit sorgfältig konstruierte Testverfahren, die einer problemadäquaten Testtheorie genügen. In den meisten Projekten ist jedoch der hierfür benötigte zusätzliche Konstruktionsaufwand zu hoch [421]. Daher sollten bei der Gebrauchstauglichkeitsbetrachtung bereits etablierte und validierte Fragebögen verwendet werden. Diese sind bei einer Verteidigung wesentlich kritikresistenter als ad-hoc-konstruierte Instrumente [421]. Sollte es dennoch nötig sein, einen Fragebogen zu konstruieren, dann sollten bestimmte Prinzipien bei der Entwicklung von Fragebögen beachtet werden, wie beispielsweise bei Bortz & Döring [36] oder Dzida et al. [128] beschrieben. Generell sollten Fragebögen nicht zu lang sein, sie sollten präzise sein und dem Wissen und Vokabular der Befragten entsprechen.

Durch ein erhöhtes Maß an Standardisierung und Antwortvorgaben können Ergebnisse besser miteinander verglichen werden. Freiere Antworten liefern mehr Informationen, jedoch sind diese schwer zu erfassen und erschweren quantitative Aussagen. Die Auswahl der Fragen (Items) muss sorgfältig erfolgen. Hierbei ist

darauf zu achten, dass keine Verzerrungen (Bias), wie beispielsweise durch Suggestivfragen entstehen [76, 349].

Der Vorteil von Fragebogenbefragungen ist, dass sie sehr flexibel sind [423], eine direkte Dokumentation und eine einfache Möglichkeit der Standardisierung bieten. Nachteilig ist, dass Anwender durch vorgegebene Antworten stärker beeinflusst werden, und Rückfragen in den meisten Fällen nicht möglich sind. Auch ist nicht auszuschließen, dass die Benutzer die Lust verlieren und keine zuverlässigen und konstruktiven Antworten mehr geben.

Die meisten Methoden können für sich allein zu keiner Produktbewertung führen, obwohl laut Dzida et al. einige Fragebögen dies fälschlicherweise aufgrund ihres Namens (z.B. Benutzerfragebogen ISONORM 9241/10 [317]) vermuten lassen [128]. Benutzerfragebögen können dazu verwendet werden, um einen ersten Eindruck aus dem Feld zu erheben, welche Nutzerprobleme als störend empfunden werden. Zusätzlich können sie als ergänzendes Tool zu weiteren Methoden eingesetzt werden.

Für die in dieser Arbeit entwickelte Standardprozedur sind subjektive Zufriedenheitsfragebögen von primärem Interesse. Diese Methoden erheben bei der Gebrauchstauglichkeitsevaluation den Anspruch, die Zufriedenheit eines Nutzers bei der Geräteverwendung und Zielerreichung zu evaluieren. Mit Hilfe dieser typischerweise kurz gehaltenen Fragebögen werden die Nutzer angehalten, ein System mit Likert-Skalen oder Semantischen Differentialen zu bewerten [36, 284]. Hierbei gibt es Fragebögen zu einzelnen Aufgaben wie den ASQ (After Scenario Questionnaire) und Fragebögen, welche die Nutzerzufriedenheit eines gesamten Systems betrachten, wie z.B. den PSSUQ (Post-Study System Usability Questionnaire), den SUS (System Usability Scale) oder den QUIS-Fragebogen (Questionnaire For User Interaction Satisfaction). Im Anhang A werden einige der wichtigsten Fragebögen der Gebrauchstauglichkeitsbetrachtung vorgestellt:

Fazit: Fragebögen stellen eine gute Methode zur Evaluation der subjektiven Benutzermeinung dar. Es gibt verschiedene standardisierte Fragebögen (vgl. Anhang A) die zu verschiedenen Phasen der Entwicklung und Evaluation angewendet werden können. Sie eignen sich besonders als ergänzende Tools für weitere Methoden.



### **Szenario Methode**

Die Szenariomethode dient der Kommunikation zwischen Experten und Anwendern. Unter einem Szenario wird im Allgemeinen ein vorstellbares Bild verstanden [41]. Im Rahmen des Usability Engineering können Szenarien in zwei Bereichen angewendet werden: als Beschreibung von Anwendungsfällen (Use Cases) und somit als Ablaufbeschreibung der Interaktion eines Anwenders mit einer Anwendung [153, 298, 349] oder als eine personalisierte Handlung, die dem Designer die Kommunikation mit dem Nutzer erleichtert. Nach Preece et al. sind demnach Szenarien eine personalisierte, fiktionale Geschichte mit Benutzern und dem Produkt innerhalb der Arbeitsumgebung [314]. Rosson & Carroll verwenden Szenarien im Zusammenhang mit dem von ihnen geprägten Begriff des Szenario-Based-Design als Methode der benutzerorientierten Entwicklung. Hier findet mit Hilfe von kleinen Geschichten die Kommunikation zwischen den Entwicklern, Auftraggebern und Benutzern eines Systems statt. Szenarien beschreiben somit die Situationen, in denen Menschen mit einem System arbeiten [345].

Fazit: Der erste Anwendungsfall ist für die Evaluation nicht geeignet, da er hauptsächlich als Ideenlieferant in der Anforderungsphase fungiert. Der zweite Anwendungsfall eignet sich sehr gut, um in einer Testsituation den Nutzer durch eine kleine Geschichte in den Nutzungskontext zu versetzen.

Eine Zusammenfassung der in diesem Kapitel beschriebenen Methoden ist im Anhang B zu finden.

### **2.4.2 Methodenvergleich**

Gemäß dem Thema dieser Arbeit, eine Standardprozedur für die Bewertung der Gebrauchstauglichkeit medizinischer Produkte zu entwickeln, ist die Auswahl geeigneter Methoden von besonderer Bedeutung. Prinzipiell ist Nielsen zuzustimmen, dass es wichtig ist, überhaupt eine Gebrauchstauglichkeitsevaluation durchzuführen, anstatt wegen der fehlenden finanziellen Möglichkeiten für das beste Verfahren, ganz auf die Evaluation zu verzichten [284]. Schließlich ist es immer noch besser, wenige Probleme zu finden als überhaupt keine. Im Fall einer Bewertung müssen jedoch Methoden ausgewählt werden, mit denen die Entdeckung der meisten Gebrauchstauglichkeitsprobleme gewährleistet ist. Dies gilt

besonders im medizinischen Bereich, in dem ein nicht entdecktes Bedienproblem eine potentielle Gefahr für Anwender und Patienten darstellt.

Das vorherige Kapitel hat gezeigt, dass es ein großes Repertoire an vielversprechenden Methoden gibt, aus dem es gilt, die richtige auszuwählen. Bisher sind in der Literatur noch keine Überlegungen für die richtige Auswahl und Kombination von Gebrauchstauglichkeitsevaluationsmethoden speziell für den medizinischen Bereich zu finden. Backhaus liefert einen Methodenvergleich und eine Methodenkombination für den Entwicklungsprozess, welche jedoch nicht auf die Evaluation zur Gebrauchstauglichkeitsbewertung übertragen werden können, da sie speziell auf den Entwicklungsbereich ausgerichtet sind [11]. Daher werden im folgenden Teil dieser Arbeit Standardmethoden der Gebrauchstauglichkeitsevaluation auf ihre Anwendbarkeit für medizinische Geräte untersucht.

Gebrauchstauglichkeitsevaluationsmethoden werden in einer Vielzahl von Studien auf ihre Wirksamkeit, ihren Nutzen bzw. ihren Mehrwert untersucht:

**Desurviers** vergleicht die Inspektionsmethoden „Heuristische Evaluation“ und „Cognitive Walkthrough“ (CWT) jeweils unter Einsatz einer Expertengruppe, einer Gruppe aus Softwareentwicklern und einer Laiengruppe mit einem Nutzertest. Insgesamt wurden bei der Gebrauchstauglichkeitsevaluation im Labor mit Nutzern 25 Probleme gefunden, darunter fünf kleine Probleme, welche leichte Irritation hervorriefen, drei mittlere Probleme, welche zu Fehlern führten und 17 schwere Probleme, die verhinderten, dass die Aufgabe gelöst werden konnte. Bei der Heuristischen Evaluation fanden Experten 80% der leichten, 29% der schweren und 67% der mittleren Probleme. Mit dem CWT konnten 67% der mittleren, aber weniger als 50% der leichten und schweren Fehler gefunden werden. Inspektionsmethoden mit Softwareentwicklern und Laien waren weitaus weniger effizient [72]. Ergebnis dieser Studie ist, dass mit Gebrauchstauglichkeits-Inspektionsmethoden Gebrauchstauglichkeitsprobleme aufgedeckt werden können. Sie sind besonders wirksam, wenn Sie von Experten angewandt werden, können jedoch Nutzertests nicht ersetzen.

**Doubleday et al.** vergleichen die Heuristische Evaluation nach Nielsen [284] mit fünf Experten der Mensch-Maschine-Kommunikation mit einem Nutzertest mit 20 Endnutzern anhand eines „Information Retrieval“-Systems. Sie zeigen dabei, dass die Probleme, welche durch diese beiden Methoden gefunden wurden, unter-

schiedlich sind. Zwar fanden die Experten mit 86 Problemen die meisten Resultate, jedoch deckten sich diese Probleme nur zu 39% (15 Probleme) mit den 38 der durch den Nutzertest gefundenen. Von den 6 mit höchstem Schweregrad bewerteten Problemen, von denen zwei unter den meist auftretenden Fehlern waren, wurden drei von den Experten nicht entdeckt und zwei nur durch einen Experten. Somit blieben den Experten offensichtlich einige der „wirklichen“, in einer Nutzungssituation auftretenden Fehler verborgen. Nach Doubleday et al. werden durch die Heuristische Evaluation Probleme entdeckt, welche gegen MMI-Prinzipien verstoßen (z.B.: fehlendes Feedback, unklare Begriffe, schlechte Sichtbarkeit). Dagegen sind Nutzertests geeignet, um Probleme zu identifizieren, welche Benutzern bei der Verwendung widerfahren. Aufgrund ihrer Ergebnisse vermuten Doubleday et al. den Unterschied zwischen den beiden Methoden darin liegend, dass mit Nutzertests die Symptome der Probleme identifiziert werden (wie z.B. die Wahl eines falschen Menüs), wogegen die Heuristische Evaluation die Ursache des Problems (z.B. schlechte Navigation als Ursache für die Wahl des falschen Menüs) und mögliche Lösungen aufdeckt. Daher kommen sie zu dem Schluss, dass für eine vollständige Durchdringung einer Benutzerschnittstelle die Verwendung von mehreren Techniken notwendig ist [120].

**Jeffries et al.** untersuchten die Effektivität der Evaluationsmethoden durch eine Heuristische Evaluation mit vier Experten, durch eine Überprüfung anhand von Software-Richtlinien, durch einen Cognitive Walkthrough (mit der Besonderheit der Durchführung in einer Gruppe und nicht bei Einzelpersonen) und durch Nutzertests mit sechs Endnutzern anhand einer Benutzerschnittstelle. Von den insgesamt 206 gefundenen Einzelproblemen wurden die meisten durch die Heuristische Evaluation entdeckt (105). Die drei restlichen Methoden waren in Bezug auf die Anzahl der gefundenen Probleme untereinander ebenbürtig. Erst nach einer Beurteilung des Schweregrades durch sieben Personen wurde ersichtlich, dass durch den Nutzertest mehr schwerwiegende Probleme auftraten. Dennoch wurden die Werte der Heuristischen Evaluation mit einem Drittel der schwerwiegendsten und zwei Dritteln der weniger schweren Probleme nicht erreicht. Jeffries et al. kommen zu der Erkenntnis, dass unter der Voraussetzung der Verwendung von Experten die Heuristische Evaluation die beste Methode ist, um bei geringem Aufwand die meisten Ergebnisse zu erhalten. Nutzertests sind gut, um gravierende Probleme zu finden, und besonders geeignet, um wiederkehrende und allge-

meine Probleme aufzudecken, jedoch unter erheblichem Aufwand und immer noch in geringerem Maß als bei der Heuristischen Evaluation. In der Studie war die Evaluation mit Richtlinien die beste der vier Techniken, um allgemeine und wiederkehrende Probleme aufzudecken. Obwohl Gutachter mit gut konzipierten Richtlinien eine breite Einsicht in ein Produkt bekommen, werden meist schwere Probleme nicht gefunden. Ergebnisse der Methode Cognitive Walkthrough sind in etwa vergleichbar mit denen der Evaluation mit Richtlinien. Eine Empfehlung der Autoren ist, dass Überprüfungen von Richtlinien und Cognitive Walkthrough eher Techniken für Designer sind, um schnell einige Probleme zu identifizieren, insbesondere, wenn keine Konsultation von Gebrauchstauglichkeitsexperten möglich ist. Mehr Vorteile bieten Nutzertests und Heuristische Evaluation. Viele der schwerwiegenden Probleme werden eher mit diesen Techniken gefunden, wobei Jeffries et al. keine eindeutige Empfehlung geben, welche der beiden Methode die bessere ist [205]. Zwar suggeriert diese Studie, dass eine Heuristische Evaluation mehr Probleme findet, jedoch ist fraglich, ob auch alle Probleme, welche Experten mutmaßlich in einem Produkt sehen, auch wirkliche Nutzerprobleme sind.

**Desurvire et al.** verglichen zwei Arten der Heuristischen Evaluation in einem Fragebogenformat (mit Experten und Laien) mit einem Nutzertest, in dem Nutzer neun Aufgaben mit dem Prototyp einer telefonbasierten Applikation zu erledigen hatten. Als Bewertungsmaß wurde in dieser Studie das Bestimmtheitsmaß "R-Quadrat"<sup>29</sup> als Indikator für Reliabilität herangezogen. Ausgehend von einer vorhergesehenen Aufgabenerfüllungsrate zeigte sich, dass die Heuristische Evaluation mit Laien ( $R^2=0,20$ ) Schwächen in Bezug auf die Reliabilität hatte. Die Expertenmethode ( $R^2=0,59$ ) und der Nutzertest ( $R^2=0,56$ ) waren hierbei in etwa gleichwertig. Allein durch den Nutzertest konnten sechs Verstöße gegen Richtlinien gefunden werden. Bei der Heuristischen Evaluation wurde dagegen nur einer gefunden.

Diese Studie zeigt, dass die Heuristische Evaluation in einem Fragebogenformat nicht geeignet ist, um spezifische Probleme einer Benutzerschnittstelle und deren Lösung zu entdecken. Gefundene Probleme decken sich meistens nicht mit de-

---

<sup>29</sup> Das Bestimmtheitsmaß  $R^2$  (R-Quadrat) ist ein Gütemaß für eine durchgeführte Regressionsanalyse. Es gibt an, wie die Beziehung (Korrelation) zwischen Werten ist.

nen in der Wirklichkeit. Heuristische Evaluationen sind allerdings effektiv, um Aufgaben zu identifizieren, in denen spezifische Probleme auftreten [74].

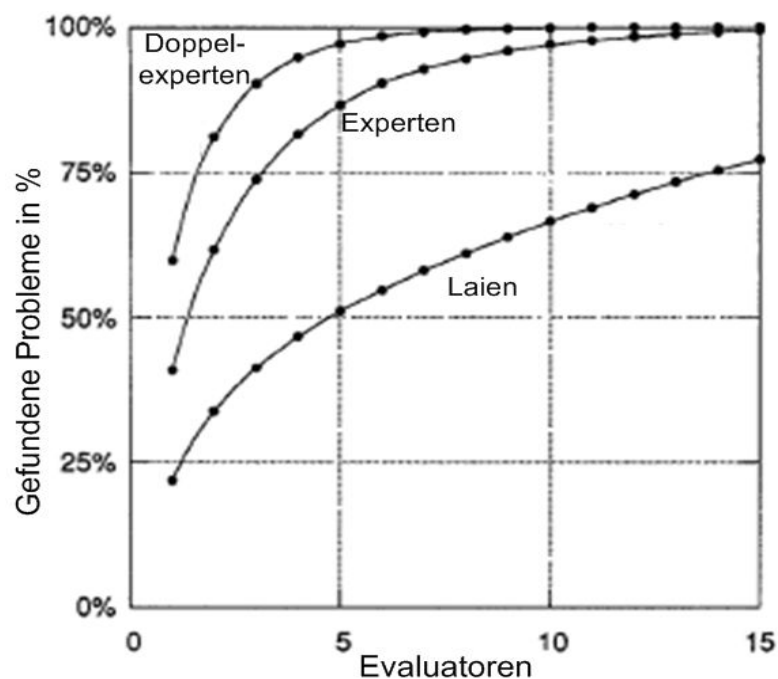
**Karat et al.** betrachteten die Effektivität eines Nutzertests mit sechs Teilnehmern sowie Einzel- und Team-Walkthrough-Methoden (2 Personen) anhand von zwei grafischen Benutzerschnittstellen. Die Walkthrough-Methoden bestanden aus einer eigenständigen Auseinandersetzung mit der grafischen Oberfläche und der Verwendung von vorgegebenen Szenarien. Bei dem Verfahren wurden 12 Gebrauchstauglichkeitsrichtlinien (Guidelines) angewendet. Die Einzel- und Teamsitzung wurde jeweils sechsmal durchgeführt.

Durch den Nutzertest konnten die meisten Probleme (159 Problemtypen für System 1 und 130 Problemtypen für System 2) gefunden werden. Die Walkthrough-Verfahren waren weitaus weniger ergiebig (Im Team: 68/54, Einzelsitzung: 49/39). Die Daten aus dieser Studie zeigen, dass der Nutzertest und der Team-Walkthrough Vorteile gegenüber dem Einzel-Walkthrough haben. Mit dem Nutzertest konnten viermal so viele signifikante Probleme gefunden werden wie beim Teamwalkthrough, und die Teams fanden 33% und 50% mehr Informationen als der Individual Walkthrough [210].

Auch nach Einbeziehung des Schweregrades der Probleme und Reduzierung auf signifikante Problembereiche wurden mit dem Nutzertest mehr als doppelt so viele schwerwiegende Probleme gefunden als mit den beiden anderen Methoden. Zu beachten ist jedoch, dass für die Teamsitzung auch eine doppelte Anzahl an Evaluatoren benötigt wird. Die Autoren lassen die Frage offen, zu welchen Ergebnissen eine gleiche Anzahl an Evaluatoren für beide Methoden führt. Weiterhin ist in dieser Studie unklar, welche Methode Karat et al. als „Walkthrough“ verstehen. Durch die Verwendung von Richtlinien handelt es sich hierbei eher um ein heuristisches Verfahren [282, 284], bei dem Probleme mit Hilfe von Gebrauchstauglichkeitsrichtlinien evaluiert werden, als um ein Walkthrough-Verfahren nach Lewis & Wharton [237] oder Bias [26].

**Nielsen** untersuchte den Effekt der Evaluatorerfahrung bei einer Heuristischen Evaluation. Im Gegensatz zur gängigen Version der Heuristischen Evaluation [290] als eine Methode, welche von Evaluatoren verwendet wird, die gewisse Erfahrungen mit Gebrauchstauglichkeitsprinzipien haben, aber nicht zwingend Gebrauchstauglichkeitsexperten sein müssen, verwendete Nielsen in dieser

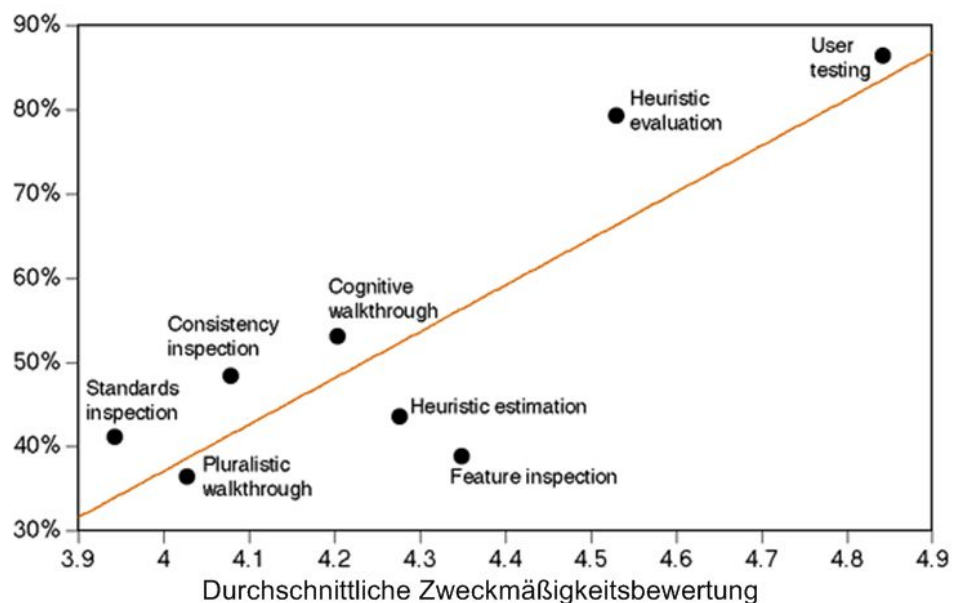
Studie drei Nutzergruppen, um mit einer Heuristischen Evaluation eine telefonbasierte Benutzerschnittstelle für einen Bank Account zu evaluieren. Die Nutzergruppen bestanden aus Laien mit keinerlei Erfahrung im Bereich der Gebrauchstauglichkeit, allgemeinen Gebrauchstauglichkeitsexperten und Doppelexperten mit Erfahrungen im Bereich der Gebrauchstauglichkeit und der zu untersuchenden Benutzerschnittstelle. Nach durchgeführter Heuristischer Evaluation und Klassifizierung der Gebrauchstauglichkeitsprobleme in leichte und schwere Probleme war erkennbar, dass die Doppelexperten die meisten Probleme fanden, gefolgt von den Gebrauchstauglichkeitsexperten. Laien fanden die wenigsten Probleme (Abb. 30). Diese Ergebnisse ergänzen vorherige Studien über die Leistungsfähigkeit einer Heuristischen Evaluation [290] um die Empfehlung, 4-6 Personen zu verwenden (vgl. Kap. 2.4.1). Bei Doppelexperten genügt eine kleine Gruppe, bei Laien hingegen muss die Gruppengröße höher angesetzt werden [282].



**Abb. 30: Anteil der gefundenen Gebrauchstauglichkeitsprobleme als Funktion der Gutachterzahl bei einer Heuristischen Evaluation**

**Nielsen** befragte in einer weiteren Studie 42 Teilnehmer eines Kurses über Gebrauchstauglichkeits-Inspektions-Methoden ein bis acht Monate nach Beendigung des Kurses bzgl. Anwendungshäufigkeit und Zweckmäßigkeit der Methoden. Die Teilnehmer konnten eine durchschnittlich 2,2-jährige Erfahrung im Gebrauchstauglichkeitsbereich aufweisen. 79% der Teilnehmer hatten bereits vor

dem Kurs Gebrauchstauglichkeitsmethoden verwendet. Sie wurden befragt, welche Methoden sie vor und nach dem Kurs angewandt hatten und wurden gebeten, den Nutzen der Methoden durch eine Bewertung einzustufen. Es zeigte sich, dass Nutzertest und Heuristische Evaluation am meisten verwendet wurden (Abb. 31). Beide Methoden wurden durchschnittlich mit zweckmäßig eingestuft, mit leicht höherem Ranking beim Nutzertest (Usefull rating=4,8) gegenüber der Heuristischen Evaluation (Usefull rating=4,5). Die weiteren Methoden wurden wesentlich weniger verwendet. Keine der Methoden wurde im Durchschnitt schlechter als 3,9 bewertet [285]. Damit zeigt die Studie, dass jede Methode dem Anwender von Nutzen sein kann und auch angewendet wird. Weiterhin wird in der Studie gezeigt, dass die meisten Anwender die Methoden für spezielle Projekte eingesetzt haben. Leider wird aber nicht klar, in welchen Phasen der Projekte sie verwendet wurden. Rückschlüsse auf die Zweckmäßigkeit der Methoden in bestimmten Produktphasen konnten daher nicht gezogen werden. Auch ist fraglich, inwiefern die Teilnehmer durch den Kurs in ihrer Methodenwahl beeinflusst waren.



**Abb. 31: Beziehung zwischen dem bewerteten Nutzen (1 = completely useless, 2 = mostly useless, 3 = neutral, 4 = somewhat useful, 5 = very useful) der einzelnen Methoden und dem prozentualen Anteil der Befragten, die die Methode verwendeten**

**Dutt, Johnson und Johnson** untersuchten die Effektivität der beiden Inspektionsmethoden Heuristische Evaluation und Cognitive Walkthrough bei der Identifizierung von Gebrauchstauglichkeitsproblemen anhand eines kommerziellen Recruitment Datenbanksystems mit einer komplexen Benutzerschnittstelle und

hoher Systemfunktionalität. Hierzu verglichen sie die beiden Methoden hinsichtlich Anzahl, Art und Schweregrad der identifizierten Fehler, des zeitlichen Aufwands und der Möglichkeit, aus den Methoden Anforderungen für ein Redesign des Produktes abzuleiten. Bei der Heuristischen Evaluation, welche an die von Nielsen und Molich [290] vorgegebene Methode angelehnt war, konnten von drei Evaluatoren (zwei mit Erfahrung aus dem Bereich der Mensch-Maschine-Kommunikation (MMK)) 37 unterschiedliche Probleme gefunden werden. Hierbei konnten die beiden Evaluatoren mit MMK-Erfahrung mehr Probleme identifizieren als der dritte. Als problematisch bei der Heuristischen Evaluation stellte sich heraus, dass die Heuristiken teilweise nicht klar abgegrenzt und missverständlich waren. Teilweise unklar war beispielsweise der Unterschied zwischen den Heuristiken „simple and natural dialog“ und „speak the users language“. Auch traten Schwierigkeiten bei der Frage auf, welche Probleme welchen Heuristiken zugeordnet werden konnten, so dass es zu Mehrfachzuordnungen kam.

Beim Cognitive Walkthrough [237], welcher mit ebenfalls drei Evaluatoren aus dem gleichen Erfahrungsbereich durchgeführt wurde, konnten 32 unterschiedliche Probleme identifiziert werden, die hauptsächlich aufgabenbezogen waren. Nach Einbeziehung des Schweregrades zeigte sich, dass die Heuristische Evaluation mehr schwerwiegende Probleme fand als der Cognitive Walkthrough. Stärken des Cognitive Walkthrough waren, dass diese Methode eine bessere Basis für ein Redesign des Produktes lieferte. Insgesamt sind Dutt et al. der Meinung, dass beide Methoden sich gegenseitig ergänzen und im Designprozess verwendet werden sollten, jedoch in unterschiedlichen Phasen [126]. Problematisch an dieser Studie ist, dass Dutt et al. eine sehr geringe Anzahl an Evaluatoren einsetzten. Zwar handelt es sich um die in der in der Literatur vorgeschriebene Anzahl [410], jedoch dürfte diese gerade für einen Methodenvergleich nicht ausreichen.

**Desurvire et al.** verglichen Nutzertest, Heuristische Evaluation und Cognitive Walkthrough miteinander. Der Nutzertest wurde mit 18 Teilnehmern durchgeführt. Heuristische Evaluation und Cognitive Walkthrough wurden jeweils mit einer Expertengruppe, einer Gruppe aus Softwareentwicklern und einer Gruppe aus Nicht-Experten mit jeweils drei Teilnehmern durchgeführt. Durch den Nutzertest wurden die meisten Probleme (25) gefunden, weitaus weniger durch die Inspektionsmethoden. Die Heuristische Evaluation mit Experten (11 Probleme = 44%)



war hierbei dem Cognitive Walkthrough mit Experten (7 Probleme = 28%) überlegen (Tab. 9). Innerhalb der Inspektionsmethoden wurden die meisten Probleme durch die Experten gefunden, gefolgt von den Softwareentwicklern und schließlich von den Nicht-Experten [73].

	Anzahl der gefundenen Probleme	Prozent
<b>Nutzertest:</b>	<b>25</b>	<b>100%</b>
<b>Heuristische Evaluation</b>		
Experten	11	44%
Softwareentwickler	4	16%
Nicht-Experten	2	8%
<b>Cognitive Walkthrough</b>		
Experten	7	28%
Softwareentwickler	4	16%
Nicht-Experten	2	8%

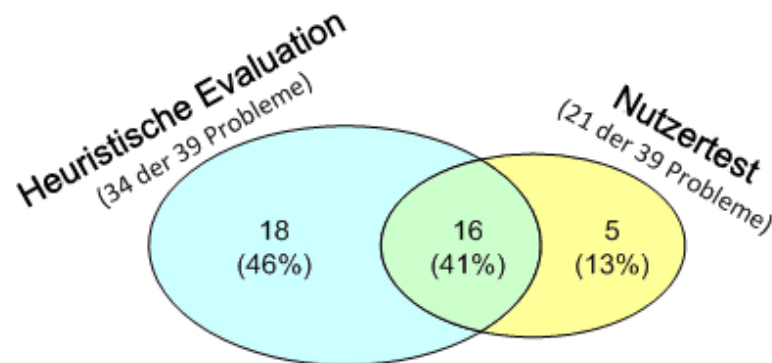
Tab. 9: Vergleich von drei Evaluationsmethoden nach Desurvire et al. [73]

Unklar bleibt bei dieser Studie, ob in der Heuristischen Evaluation und im Cognitive Walkthrough Probleme gefunden wurden, die nicht in dem Nutzertest auftauchen. Desurvire et al. gehen davon aus, dass es sich nur bei Problemen, welche im Nutzertest gefunden wurden, um wirkliche Probleme handelt. Schwer interpretierbar sind die Ergebnisse aufgrund der unterschiedlichen Teilnehmerzahl, denn 18 Teilnehmern aus einem Nutzertest stehen drei Evaluatoren bei den Inspektionsmethoden gegenüber. Die Unterschiede der Ergebnisse zwischen dem Erfahrungsgrad der Evaluatoren entsprechen der Studie von Nielsen [290].

**Fu et al.** untersuchten eine Heuristische Evaluation und einen Nutzertest anhand der Fehlerkategorisierung von Gebrauchstauglichkeitsproblemen nach den drei Ebenen der Handlungsregulation von Rasmussen (vgl. Kap. 1.2.1.1). Dementsprechend stehen Gebrauchstauglichkeitsprobleme in starkem Zusammenhang mit dem fähigkeits-, regel- und wissensbasierten Leistungsniveau. Fu et al. führten bei einer internetbasierten Multimediaapplikation einen Nutzertest mit acht typischen Nutzeraufgaben und eine Heuristische Evaluation durch. Letztere folgte keinen besonderen Regeln, wie z.B. bei Nielsen [284]. Ergonomie-Experten wurden die Aufgaben für den Nutzertest vorgelegt, und ihnen wurden 1,5 Stunden Zeit gegeben, um Probleme zu identifizieren.

Insgesamt wurden 130 Probleme durch beide Methoden identifiziert. Nach Entfernung der Doppelnennungen blieben 39 unterschiedliche Probleme übrig (Abb. 32). Ergebnis ist, dass die Heuristische Evaluation auf der Ebene von Gewohnheiten, Fertigkeiten und Regeln am besten abschneidet und Nutzertests auf Wis-

sensebenen erfolgreicher sind [142]. Beide Methoden konnten insgesamt 16 Probleme gemeinsam evaluieren, die Heuristische Evaluation allein weitere 18 und der Nutzertest weitere fünf. Aus den Ergebnissen dieser Studie erkennt man, dass beide Methoden gemeinsam verwendet werden sollten, um eine maximale Anzahl von Gebrauchstauglichkeitsproblemen zu identifizieren. Bei der Inspektionsmethode finden die Experten mehr Probleme. Unklar bleibt jedoch in dieser Studie, wie schwerwiegend die gefundenen Probleme waren.



**Abb. 32: Gefundene Probleme durch eine Heuristische Evaluation und einen Nutzertest**

Bei den vorgestellten Untersuchungsergebnissen (Zusammenfassung in Anhang C) erkennt man teilweise eine Inkonsistenz. Generell ist feststellbar, dass Nutzertest und Heuristische Evaluation den anderen Methoden bei der Auffindung von Gebrauchstauglichkeitsproblemen überlegen sind [73, 142, 205, 285]. Strittig ist, welche der beiden Methoden den größeren Nutzen stiftet. Während beispielsweise die Studien von Jeffries et al. [205] und Fu et al. [142] die Heuristische Evaluation als die ergiebiger Methode erscheinen lassen, fallen die Studien von Desurvire [72, 73, 74], Karat et al. [210] und Nielsen [285] zugunsten des Nutzertests aus. Die unterschiedlichen Ergebnisse können durch den Schweregrad und die Art der gefundenen Fehler erklärt werden. Nach Doubleday findet zwar die Heuristische Evaluation mehr Probleme, jedoch werden mit dem Nutzertest die schwerwiegendsten Probleme aufgedeckt [120]. Grundsätzlich sprechen die Ergebnisse der verglichenen Studien für eine Methodenkombination, durch welche die spezifischen Stärken einzelner Methoden vereint werden. Um also eine Benutzerschnittstelle möglichst im Rahmen einer summativen Evaluation vollständig zu durchdringen und die meisten Gebrauchstauglichkeitsprobleme aufzudecken, sollten Heuristische Evaluation und Nutzertest zusammen verwendet werden [12, 120, 124, 142, 184, 270, 284, 372]. Die restlichen Methoden sind allerdings nicht wertlos, sie haben nur einen anderen Anwendungsrahmen. Sie

sind eher in früheren Stadien des Designprozesses anzuwenden, wenn ein Nutzertest nicht möglich ist, oder sie dienen als "discount methods", wenn wenig monetäre Mittel, Zeit oder Experten vorhanden sind [204].

Allgemein ist feststellbar, dass bei geringem Kosteneinsatz und beschränktem zeitlichen Aufwand Inspektionsmethoden zu bevorzugen sind. Sollte jedoch Wert auf Genauigkeit, Reliabilität und Validität gelegt werden, können empirische Methoden nicht ersetzt werden. Verschiedene Studien belegen, dass viele der „Probleme“, welche bei einer Heuristischen Evaluation gefunden werden, keine wirklichen Probleme sind [12, 13, 120]. Bailey, Allen und Raillo [13] überprüften bei Studien von Nielsen und Molich [290] die dort genannten Ergebnisse. An dem gleichen System führten sie einen Nutzertest durch. Anhand von Leistungsmaßen (Aufgabenzeit, Zielerreichung und Fehleranzahl), Beobachtungen und Nutzerkommentaren wurden die schwerwiegendsten Probleme identifiziert. Mit Hilfe der Ergebnisse wurde die Benutzerschnittstelle angepasst und erneut evaluiert. Bereits nach zwei Iterationen war keine Verbesserung der Leistungs- und Zufriedenheitsmaße feststellbar. Aus dem Vergleich der behobenen Probleme mit denen der Heuristischen Evaluation schließen Baileys et al., dass lediglich zwei von den 29 durch die Inspektionsmethode gefundenen Probleme relevant waren und die weiteren 27 als nicht ausschlaggebend angesehen werden konnten. Die Studie zeigt, dass nicht jedes mit einer Inspektionsmethode entdeckte Problem ein wirkliches Problem der Schnittstelle darstellt. Daher wird von vielen Experten empfohlen, dass ein gut durchdachter Nutzertest einer Heuristischen Evaluation folgen sollte [12, 120, 142, 284]. Durch diese Reihenfolge können die in einer Heuristischen Evaluation vermuteten Probleme durch einen Nutzertest mit realen Nutzern auf ihre Relevanz und ihren Wahrheitsgrad mit Hilfe geeigneter Szenarien überprüft und verifiziert und zusätzlich weitere Probleme aufgedeckt werden.

Bei der Fragestellung dieser Arbeit, die Gebrauchstauglichkeit medizinischer Geräte zu bewerten, ist eine möglichst genaue Durchdringung der Benutzerschnittstelle besonders wichtig. Auch laut der Deutschen Akkreditierungsstelle Technik (DATech) eignen sich Inspektionsverfahren nur bedingt zu einer Konformitätsprüfung bzgl. der für Gebrauchstauglichkeit relevanten Normen. Mit ihnen kann zwar die Effektivität (vgl. Kap 2.4) eines Systems getestet werden, jedoch nicht die Effizienz und die Nutzerzufriedenheit. Dies lässt sich nur unter Einbeziehung der richtigen Nutzer des Systems erreichen [76]. Im Gegensatz zu den Inspektionsmethoden bieten Nutzertests einen potentiell hohen Validitätsgrad, da

neben aufgaben- und systemspezifischen Charakteristika auch komplexe umgebungs- und benutzerspezifische Einflussfaktoren (wie z.B. Stress, Müdigkeit, Aufmerksamkeit, Temperatur, Klima) durch Videoanalysen, Observation, Interviews etc. innerhalb von Labor- und Feldversuchen abgebildet werden können [202]. Weiterhin können durch Nutzertests die Gebrauchstauglichkeitsmaße Effektivität, Effizienz und Zufriedenheit evaluiert, Probleme aufgedeckt und Informationen zu deren Auftretenshäufigkeit und Schwere gewonnen werden [385].

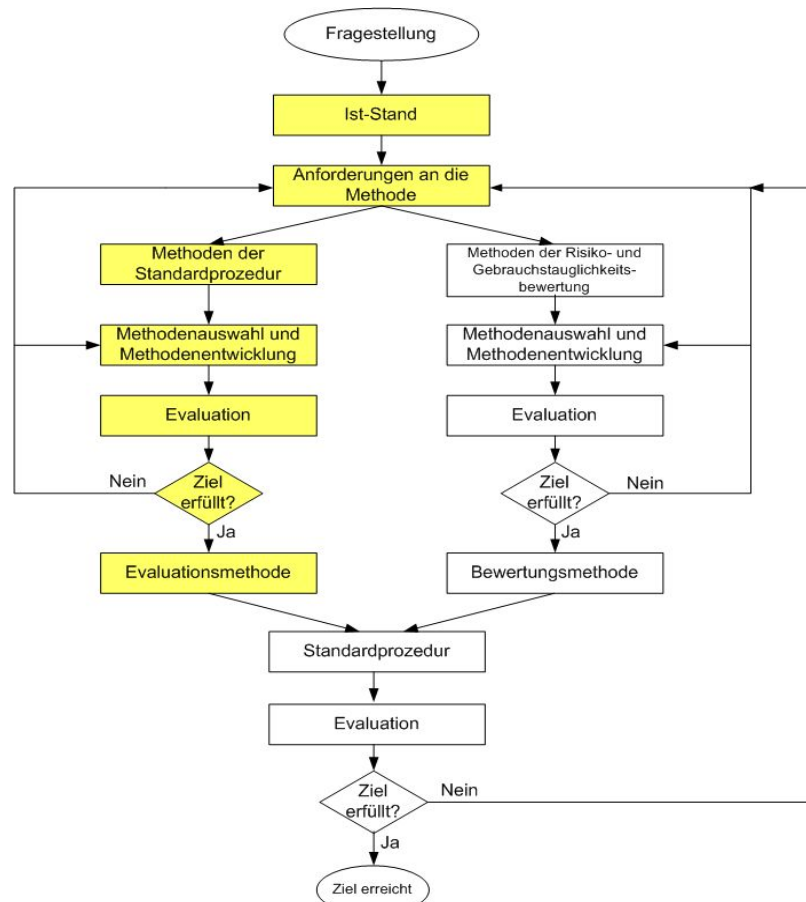
Für den medizinischen Bereich ist neben den Gebrauchstauglichkeitsmaßen, welche durch die beschriebenen Methoden evaluiert werden können, auch die Sicherheit der Anwender und der Patienten wichtig (vgl. Kap. 2.2.2.2). Gerade weil medizinische Geräte sicherheitsrelevante Systeme sind, ist es von besonderer Bedeutung, möglichst alle Gebrauchstauglichkeitsprobleme in einer Evaluation zu finden. Dies ist nur mit empirischen Verfahren unter Einbindung der Nutzer in die Evaluation möglich. Sie untersuchen, im Gegensatz zu formalen oder heuristischen Evaluationsverfahren, die tatsächliche Interaktion der Geräte mit dem Nutzer, anstatt sich auf theoretische Modelle zu beziehen. Auch können mit diesen Methoden „worst case“-Szenarien und spezielle Stress und Anwendungssituationen überprüft werden [187]. Die bisherigen Methoden erfüllen den Anspruch der Sicherheit nicht, mit ihnen können lediglich Gebrauchstauglichkeitsprobleme evaluiert werden. Der Nutzertest identifiziert zwar, im Gegensatz zu anderen Evaluationsmethoden, Probleme bei der Geräteverwendung und gibt Hinweise über deren Schweregrad [372], jedoch kann kein spezieller Bezug zu einer damit verbundenen möglichen Patientengefährdung hergestellt werden. Im weiteren Verlauf soll eine Methode entwickelt werden, mit der die evaluierten Probleme hinsichtlich ihrer sicherheitskritischen Relevanz bewertet werden können.

### 3. Entwicklung einer Methode zur Überprüfung der Gebrauchstauglichkeit medizinischer Produkte

#### 3.1 Ziele und Vorgehensweise

Um die Evaluationsmethode für Gebrauchstauglichkeitsprobleme in der Medizin zu entwickeln, werden eine Methode zur Evaluation der Gebrauchstauglichkeit und eine Methode zur Bewertung der Gebrauchstauglichkeit zuerst getrennt voneinander ausgearbeitet und evaluiert. Nach erfolgreicher Evaluation werden die beiden Methoden zusammengeführt und als eine Methode evaluiert.

In diesem Kapitel wird der Ist-Stand der Gebrauchstauglichkeitsentwicklung aufgezeigt und die allgemeinen Anforderungen an eine Evaluations- und Bewertungsmethode aufgestellt. In den folgenden Kapiteln werden, ausgehend von den aufgestellten Anforderungen, in einem iterativen Prozess, mit wiederkehrenden Entwicklungs- und Evaluationsphasen (Abb. 33), die Methoden konkretisiert und in eine Standardprozedur eingearbeitet.



**Abb. 33: Entwicklungsprozess der Standardprozedur. Der in diesem Kapitel behandelte Evaluationsteil ist farbig markiert.**

## **3.2 Ist-Situation: Diskussion der bestehenden Prüfverfahren im Hinblick auf die Standardprozedur**

### **3.2.1 Ergonomieprüfung in der Medizin**

Zur Ermittlung der Gebrauchstauglichkeit von Medizinprodukten gibt die Norm DIN EN IEC 60601-1-6 keine speziellen Methoden vor. In Praxis und Theorie lassen sich verschiedene Methoden finden, mit dem Potential, die Gebrauchstauglichkeit medizinischer Produkte zu messen. Im Folgenden werden mögliche Methoden vorgestellt und diskutiert, um hieraus Erkenntnisse für eine Standardprozedur abzuleiten.

#### **Probestellung**

Eine Probestellung ist in den meisten Krankenhäusern und Arztpraxen die gängigste Methode, um Medizinprodukte zu testen. Eine Umfrage auf zwei Konferenzen von Matern et al., an denen 98 Ingenieure und Planer von Medizintechnik sowie medizinisches Personal teilnahmen, zeigt, dass in über 80% der Fälle eine Probestellung entscheidend für die Beschaffung von OP-Geräten ist [257]. Probestellungen laufen in den meisten Fällen nicht nach einem speziellen Schema oder nach einem Leitfaden ab. Die Produkte werden in der Regel lediglich oberflächlich geprüft, wobei Gebrauchstauglichkeitsgesichtspunkte in den seltensten Fällen Beachtung finden. Es kann davon ausgegangen werden, dass in den meisten Kliniken das ergonomische Fachwissen fehlt, um medizinische Produkte hinsichtlich ihrer Gebrauchstauglichkeit zu überprüfen. Zwar werden öfter verschiedene Geräte in einer Probestellung verglichen, jedoch erfolgen diese Prüfungen weder nach einem konkreten Schema noch mit der nötigen Intensität, sodass aussagekräftige Ergebnisse in den meisten Fällen nicht erreicht werden können. Schon gar nicht werden bei den Probestellungen im Routinebetrieb kritische Zustände, wie z.B. Stress- und Notfallsituationen bei der Bedienung überprüft. Derartige Ereignisse treten aber bei kritischen Kombinationen von Einflussfaktoren immer wieder auf, sodass mögliche verborgene Bedienprobleme zu einem erhöhten Risiko führen können. In verschiedenen Ländern (z.B. Kanada, Schweden, Niederlande) werden daher bereits anstelle einer Probestellung Gebrauchstauglichkeitstests durchgeführt [187].

### **Use Error Checklist**

Die FDA veröffentlicht auf ihrer Homepage eine "Use Error Checklist"<sup>30</sup>, um Bedienfehler medizinischer Geräte zu entdecken und zu verstehen. Die Checkliste besteht aus 18 Fragen zur Evaluation einer allgemeinen Bedienbarkeit und zu Sicherheitsmängeln. Eine weitere Checkliste mit 17 Fragen, welche als Entscheidungshilfe für den Anwender bei der Anschaffung medizinischer Geräte dienlich sein könnte, ist ebenfalls auf der Internetseite der FDA erhältlich<sup>31</sup>. Diese Liste prüft mögliche Einschränkungen des Anwenders, welche z.B. die Bedienung des Gerätes beeinflussen können. Des Weiteren wird die Anwendbarkeit des Gerätes in seiner Arbeitsumgebung geprüft sowie die eventuelle Beeinflussung der Bedienung durch Gerätecharakteristika. Die Betrachtungsweise beider Checklisten ist allerdings als oberflächlich zu bezeichnen. Dem Evaluator bleiben genaue Informationen bei der Nutzung verborgen, sodass die umfassende Bewertung eines medizinischen Gerätes nicht möglich ist. Gebrauchstauglichkeitsprobleme können nur in geringem Maße entdeckt werden. Die Checkliste könnte maximal als Teilaspekt einer Experteninspektion (vgl. Kap. 2.4.1) dienen, um im Vorfeld einer Prüfung potentielle Probleme zu identifizieren und in einem Gebrauchstauglichkeitstest zu verifizieren.

### **Framework nach Lowery & Martin**

Das von Lowery & Martin erarbeitete Rahmenwerk zur „Evaluation von Softwaresystemen der Gesundheitsfürsorge aus der Perspektive der Gebrauchstauglichkeit“ soll die wichtigsten Gebrauchstauglichkeitskriterien bei einer Krankenhaussoftware zusammenfassen. Durch Literaturrecherche der Gesundheitsfürsorge und der Gebrauchstauglichkeitsfachliteratur entwickelten sie sechs Hauptpunkte, die bei einer Gebrauchstauglichkeitsbetrachtung für derartige Systeme wichtig sind: 1) Logische Organisation der Verfahren, 2) Screen-Design für die Dateneingabe, 3) Fehlerbehandlung, 4) Datenbeschaffung/Berichterstellung, 5) Lernen/Hilfe, 6) Konsistenz [247].

Das auf Software beschränkte Rahmenwerk gibt zusätzlich Evaluationskriterien und zu beachtende Hinweise im Sinne von Anforderungen an medizinische Softwareprodukte, lässt jedoch genaue Bewertungs- oder Erhebungsmethoden vermissen. Auch eine Bewertung der Gebrauchstauglichkeit eines medizinischen

---

<sup>30</sup> [www.fda.gov/cdrh/humfac/useerrorchecklist.pdf](http://www.fda.gov/cdrh/humfac/useerrorchecklist.pdf) (27.10.2008)

<sup>31</sup> [www.fda.gov/cdrh/humfac/you\\_choose\\_checklist.pdf](http://www.fda.gov/cdrh/humfac/you_choose_checklist.pdf) (27.10.2008)

Softwareproduktes wird mit dem Rahmenwerk nicht möglich sein. Dennoch können die Hinweise in die Gebrauchstauglichkeitsprüfung eingebettet werden, um hieraus Bewertungskriterien zu entwickeln.

### **MECHA**

Die von Fields, Paternò et al. entwickelte Methode „MECHA“ (Method for Evaluation of Cooperation, Hazards, and Allocation) soll Gebrauchstauglichkeit in sicherheitskritischen Systemen evaluieren [140]. Hierzu sollen Entwicklern Elemente für eine integrierte, qualitative und strukturierte Analyse von Gebrauchstauglichkeit und Sicherheitsaspekten bereitgestellt werden. Die Methode hat das Ziel, einige dieser Probleme durch die Verwendung dreier Hauptbestandteile zu beheben: 1. Aufgabenmodellierung und Identifizierung der wichtigsten Anwenderszenarien für die Benutzung, 2. Analysieren möglicher Abweichungen vom „normalen“ Use Case und 3. Berücksichtigung des Zusammenwirkens zwischen verschiedenen Anwendern und dem System. Als Abweichung ist hierbei das Verhalten eines Benutzers oder des Systems zu sehen, welches einem erwarteten Verhalten widerspricht. Diese Abweichungen können beispielsweise auch auf eine Fehlbedienung aufgrund eines Gebrauchstauglichkeitsproblems zurückzuführen sein.

Die Methode stützt sich darauf, dass Sicherheit nicht eine Eigenschaft einer speziellen Handlung oder Aufgabe ist, sondern sich in den Zusammenhängen eines ganzen Systems widerspiegelt. Im Gegensatz zu gängigen Methoden, wie z.B. der Heuristischen Evaluation (Kap. 2.4.1.1), wird bei MECHA zusätzlich und ausdrücklich das Zusammenwirken verschiedener Elemente des Systems und von Gefahren berücksichtigt, um Entwicklern eine Evaluation verschiedener Kommunikationstechnologien in komplexen, sicherheitskritischen Systemen zu ermöglichen.

Die Methode ist in vier Phasen unterteilt. In der ersten Phase sind eine Aufgabenanalyse und eine Modellierung des derzeitigen Systems und der möglichen Designoptionen durchzuführen. In der zweiten Phase werden spezifische Sequenzen von Aufgaben und entsprechende Szenarien ausgewählt. Diese müssen, wenn es sich um einen sicherheitskritischen Kontext handelt, so gestaltet sein, dass sie auf Situationen fokussieren, die ein System in einen gefährlichen Zustand bringen können. In der anschließenden dritten Phase werden mögliche Risiken und Abweichungen nach vordefinierten Kategorien analysiert. Grundlage hierfür ist eine



von Paternò et al. entwickelte Methode [304], welche auf der Risikoanalysemethode HAZOP (► Kap. 5.4.1) [380, 291] aufbaut. Ergänzend zu HAZOP analysiert ihre Methode Abweichungen in sicherheitskritischen Systemen anhand von Aufgabenmodellen, sodass auch die Kommunikation zwischen den Systemkomponenten betrachtet wird. Spezielle Fragen dieser Phase der MECHA Methode sind:

- Welches sind die möglichen Gefahren, die als Folge von Abweichungen, Störungen in der Kommunikation, fehlerhaften Handlungen auftreten können?
- Gibt es Faktoren, die dazu neigen, fehlerhafte Kommunikation, fehlerhafte Handlungen, falsche Bewertungen hervorzurufen?
- Welche Empfehlungen an die Entwickler der Mensch-Maschine-Schnittstellen können zur Milderung gefährlicher Zustände und Auswirkungen führen?

In der letzten Phase werden die Koordinationsanforderungen analysiert, welche zwischen den Teilnehmern verlangt werden. Es wird untersucht, inwieweit das Design diese Koordinationsanforderungen unterstützt [140].

Durch die MECHA-Methode können Risiken in sicherheitskritischen Systemen aufgedeckt werden. Hierbei ist es unerheblich, dass die Methode nicht speziell auf den medizinischen Kontext ausgerichtet ist. Bei Beachtung der speziellen Rahmenbedingungen dieses Arbeitssystems wird eine Übertragung hierauf möglich sein, besonders im Hinblick darauf, dass diese Methode eine Evaluation anhand von Szenarien durchführt und keine Nutzertests oder Beobachtungen vorsieht. Hierin liegt allerdings auch der Nachteil dieser Methode in Anbetracht der in dieser Arbeit zu entwickelnden Bewertungsmethodik. Durch die Auswahl von Szenarien, die auf Situationen fokussieren, welche ein System in eine gefährliche Situation bringen können, werden nur extreme Situationen erfasst und nicht der gänzliche Nutzungsrahmen, wodurch eine objektive Bewertung des Produktes nicht möglich ist. Auch andere Aufgaben müssen betrachtet werden, um eine allgemeine Gebrauchstauglichkeit bewerten zu können. Des Weiteren erfordert die MECHA-Methode einen erheblichen Aufwand von den Evaluatoren, denn sie verlangt, dass verschiedene Aktivitäten durchgeführt werden: Aufgaben-Analyse, Identifizierung von Gestaltungsmöglichkeiten, Analyse der Abweichungen sowie Zusammenstellung einer Reihe von sinnvollen Szenarien. Dennoch rechnet sich dieser Aufwand, wenn hierdurch die Sicherheit erhöht und vielleicht Menschenleben gerettet werden.

### **DATEch Prüfverfahren**

Das Prüfhandbuch „Leitfaden für die ergonomische Evaluierung von interaktiven Systemen auf Grundlage von DIN EN 9241, Teile 11 und 110“ der Deutschen Akkreditierungsstelle Technik e.V. (DATEch) [77] enthält eine Prüfanleitung auf Normkonformität mit der zugrunde gelegten Norm<sup>32</sup>. Das Verfahren soll Herstellern ermöglichen, ihre Produkte bezüglich Gebrauchstauglichkeit zu bewerten und Verbesserungen abzuleiten. Laut Prüfhandbuch werden zur Prüfung keine speziellen Laboreinrichtungen benötigt, da sie im tatsächlichen Nutzungskontext durchgeführt wird. Im Gegensatz zu den Prüfungen nach den Normen DIN EN 9241 Teile 11 bis 17 und 110 [112], die eine vollständige Verifikation<sup>33</sup> für den Nachweis der Normkonformität fordern, soll der Prüfaufwand nach dem DATEch Prüfhandbuch relativ gering sein, da lediglich die aus den Teilen 11 und 110 abgeleiteten Prüfkriterien verifiziert werden (Kap. 2.4). Die Prüfung sieht vor, für jedes Kriterium eine Konformitätsannahme zu belegen. Sollte sich ein Kriterium in der Verifikation nicht behaupten, so sieht diese Prüfung eine Erhärtungsprüfung zur Ermittlung der Bedeutsamkeit der Normverletzung vor. Hierzu werden dann entsprechend die Teile 11-17 der DIN EN 9241 herangezogen. Entwickler können hieraus Vorschläge zur normkonformen Verbesserung des Produktes entnehmen. Die Prüfung ist auf der Grundlage der in der DIN EN 9241-11 definierten Prüfkriterien Effektivität, Effizienz und Zufriedenstellung aufgebaut und in drei Teile gegliedert. Der erste Teil dient der Prüfvorbereitung. Zuerst müssen der Nutzungskontext und die Aufgaben analysiert werden, um daraus Nutzungsanforderungen abzuleiten, aus denen in einem späteren Schritt Prüfkriterien abgeleitet werden. Hierzu werden episodische Beschreibungen von Aufgaben und Tätigkeiten in ihrem Kontext (Kontextszenarien) erhoben. In der nächsten Phase wird die Prüfung durchgeführt. Dafür muss für jedes Prüfkriterium das passende Produktmerkmal ausgewählt, resp. die geforderte ausgeübte Tätigkeit identifiziert und dann auf Normkonformität überprüft werden. Als Methoden zur Durchführung der Teile 1 und 2 schlägt das Prüfhandbuch eine Aufgabenanalyse, eine Inspektion, eine teilnehmende Beobachtung, eine Benutzerbefragung oder eine Dokumentenanalyse vor. In der Regel reicht eine Inspektion des Produktes aus, um Merkmalsübereinstimmungen festzustellen, wenn das Prüfkriterium als Merkmal defi-

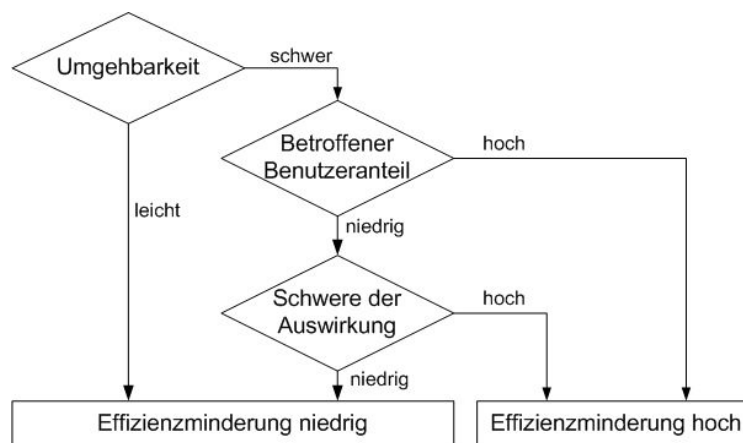
---

<sup>32</sup> Für die methodische Umsetzung des Entwicklungsprozesses gemäß ISO 13407 stellt DATEch ein weiteres Prüfhandbuch für den Usability-Engineering-Prozess bereit [75].

<sup>33</sup> Verifikation: Vorgang, durch den ein vermuteter oder behaupteter Sachverhalt als wahr angesehen werden kann.

niert ist. Ist das Prüfkriterium als Tätigkeit definiert, so ist die teilnehmende Beobachtung die zu bevorzugende Methode. Bei aufgefallenen Normabweichungen wird eine Erhärtungsprüfung durchgeführt.

Nach durchgeführter Prüfung müssen die vermuteten Normabweichungen mit Hilfe einer Entscheidungstabelle bewertet werden. Die Einstufung der Effektivitätsminderung, Effizienz-minderung und Minderung der Zufriedenstellung wird mit hoch oder niedrig vorgenommen. Als Hilfestellung für die Ermittlung dienen Entscheidungsbäume (Abb. 34).



**Abb. 34: Entscheidungsbaum für Effizienz-minderung [77, 128]**

Das DATech Prüfverfahren ist an die DIN EN 9241-11 angelehnt. Es ist in erster Linie für Hersteller konzipiert worden und ohne spezielle Vorbildung anwendbar. Es handelt sich um ein Prüfverfahren, das für die Festlegung von Normabweichungen in interaktiven Systemen entwickelt wurde. Mit diesem Verfahren werden die Mindestanforderungen der Normen überprüft. Für eine ausreichende Bewertung der Gebrauchstauglichkeit ist dieses Verfahren nicht tiefgreifend genug, denn es werden nur die Mindestanforderungen bewertet. Das Verfahren ist auch nicht speziell auf sicherheitskritische Geräte ausgelegt und berücksichtigt daher nicht die für medizinische Geräte relevanten Sicherheitsanforderungen.

### **KAN 31-Fragenkatalog zur Überprüfung sicherheitskritischer Eigenschaften**

Die Kommission für Arbeitsschutz und Normung (KAN) veröffentlicht in ihrem „Ergonomieleitfaden für die Gestaltung von Medizinprodukten“ [186] einen Fragenkatalog zur Identifizierung von Gefährdungen mit 18 Fragen, welche sicherheitsrelevante Faktoren von Medizinprodukten betreffen. Der Fragebogen stellt hierbei keinen Evaluationsfragebogen dar, sondern dient eher als Gedankenstütze, um durch beispielhafte Fragen den Fokus des Entwicklers auf den Anwender

zu schärfen. Die Fragen betreffen eher die Risikobetrachtung. Die Gebrauchstauglichkeit wird hierbei lediglich als Teil der Identifizierung von Gefährdungen verstanden.

#### **Schema nach Hölscher et al.**

Hölscher et al. empfehlen eine Expertenbegutachtung und Einbeziehung der Nutzer für die Gebrauchstauglichkeitsprüfung von Medizinprodukten [187]. Zur Prüfung werden sogenannte „worst case“-Szenarien in simulierter Umgebung eingesetzt, um komplexe Situationen mit Störungen und Wechselwirkungen effektiv zu testen. Es wird jedoch keine genaue Methode vorgegeben. Je nach Komplexität des Medizinproduktes kann eine Prüfung durch einen Test, eine Inspektion oder eine Abfrage von Erfahrungen evaluiert werden. Zur Unterstützung der Evaluation und zur Analyse eines ergonomischen Arbeitssystems haben Hölscher et al. ein Schema entwickelt, welches „den am Entstehungsprozess des Medizinproduktes beteiligten Personen die wichtigsten Felder und Gesichtspunkte aufzeigt, die für eine gebrauchstaugliche Gestaltung überdacht und eventuell berücksichtigt werden müssen“.

Das Schema kann herangezogen werden, um Prüfkriterien für die Gebrauchstauglichkeitsprüfung eines medizinischen Gerätes abzuleiten. Hierzu müssen die entsprechenden relevanten Prüfkriterien ausgewählt und operationalisiert werden. Das sehr detaillierte Schema ist eine sinnvolle Hilfestellung bei der Entwicklung von Prüfkriterien. Es lässt dem Evaluator viele Freiheitsgrade bei der Prüfung und schreibt keinen speziellen Prüfprozess vor. Für eine Bewertung der Gebrauchstauglichkeit ist es ungeeignet, da vergleichbare Ergebnisse hierbei schwierig zu erhalten sind. Das Schema ist eher als Methode zur Entdeckung von Problemen und als Hilfestellung für die Entwickler medizinischer Geräte zu sehen, da durch eine Fülle an Prüfpunkten viele Bereiche abgedeckt werden.

#### **PROMEDIKS – Prozessorientierte Medizintechnik in klinischen Systemen**

Die von Backhaus erarbeitete PROMEDIKS-Methode [11] ist eine speziell auf die Medizintechnik abgestimmte Methode, welche die Einflussfaktoren Bedienbarkeit (Gebrauchstauglichkeit) und Funktionalität miteinander verbindet. Diese Methode wurde aus bestehenden Lösungsansätzen zur Prozessunterstützung und zur Bedienbarkeit entwickelt. Es handelt sich hierbei um eine Kombination der TOPICS-Methodik (Together Optimizing Processes in Clinical Systems) [253] und

der Gebrauchstauglichkeitsevaluationsmethoden (Benutzerbefragung, Expertenbefragung und Nutzertest (vgl. Kap. 2.4.1)), mit der Aufgabe, bei der Entwicklung, Auswahl und Implementierung von Medizintechnik zu unterstützen. Die PROMEDIKS-Methodik gliedert sich in drei Phasen:

1. Analyse und Bewertung der funktionalen Prozessunterstützung, bestehend aus:
  - Partizipativer Prozessanalyse
  - Prozessvisualisierung
  - Ermitteln der Prozessunterstützung
2. Analyse und Bewertung der Gebrauchstauglichkeit
  - Ermitteln von Anwendungsszenarien
  - Auswahl der Gebrauchstauglichkeitsmethode(n)
  - Gebrauchstauglichkeitsevaluation
3. Synthese der Ergebnisse, wahlweise qualitativ oder quantitativ

Die TOPICS-Methode wurde zur Analyse und Optimierung von komplexen Prozessen im Arbeitssystem Krankenhaus entwickelt [253]. Die Methode gliedert sich in die Abschnitte partizipative Prozessflussvisualisierung und problemspezifischer Prozessmodulvergleich. In der ersten Phase werden die im Untersuchungsfeld stattfindenden Arbeitsprozesse visualisiert, verifiziert, analysiert und bewertet. Die Datengewinnung erfolgt mit einer offenen, nicht teilnehmenden Beobachtung [36]. In einem späteren Verlauf werden mittels eines einfachen Symbolvorrats Prozessflüsse visuell dargestellt. Zusätzlich werden von dem Versuchsleiter beobachtete oder durch Befragungen evaluierte Stärken und Schwächen des Prozesses in die Prozessflussdiagramme übernommen, um schließlich quantifizierbare Kenngrößen zu ermitteln, die einen Vergleich einzelner Prozessmodule unterschiedlicher Prozessflüsse ermöglichen sollen. Durch diese Ergebnisse können potentielle Engpässe und Schwachstellen in der Prozessflussstruktur erkannt und optimiert werden. In der zweiten Phase, dem problemspezifischen Prozessmodulvergleich, werden Schritte durchgeführt, um die ermittelten problematischen Prozessmodule zu verbessern. Bereits analysierte Arbeitssysteme können hierbei geeignete Lösungsvarianten bieten. Marsolek stellt hierfür eine Datenbank zur Verfügung, die das Identifizieren und Auswählen von alternativen Prozessmodulen unterschiedlicher Arbeitssysteme unterstützt [11, 253].

Als Methoden des Usability Engineerings zur Ermittlung der Bedienbarkeit der Produkte schlägt Backhaus eine Kombination aus den Methoden Benutzerbefra-

gung, Expertenbefragung und Nutzertests vor. Diese sollen entsprechend dem Stand der Produktentwicklung und dem erwarteten Bedienrisiko sinnvoll zur Evaluation gewählt werden. Zur Ermittlung der Anwendungsszenarien für die Gebrauchstauglichkeitsevaluation liefert die TOPICS-Methode die für die Gebrauchstauglichkeit relevanten Prozessmodule. Neben Prozessmodulen repräsentativer Standardanwendungen sollen auch solche ausgewählt werden, in denen ein erhöhtes Bedienrisiko vorliegt.

Die von Backhaus entwickelte Methode untersucht Gebrauchstauglichkeit und Funktionalität medizinischer Geräte im medizinischen Kontext. Die Methode kombiniert in gelungener Weise Lösungsansätze der Prozessunterstützung und der Gebrauchstauglichkeit, um Gebrauchstauglichkeitsprobleme zu analysieren. Die Methode ist vor der Entstehung der DIN EN IEC 60601-1-6 und der DIN EN IEC 62366 entwickelt worden, sodass eine Ausrichtung speziell auf die Anforderungen dieser Normen nicht durchgeführt werden konnte. Da die Methode primär darauf zugeschnitten ist, entwicklungsbegleitend Gebrauchstauglichkeitsprobleme aufzudecken, werden Risikoaspekte und eine Bewertung der gefundenen Probleme nur am Rande betrachtet und keine speziellen Vorgehensweisen für Bewertungen von Gebrauchstauglichkeitsproblemen unter dem Aspekt eines Risikomanagements vorgeschlagen. Um Geräte zu vergleichen und bewerten zu können, muss eine einheitliche Vorgehensweise durchgeführt werden. Da die PROMEDIKS-Methode eher ihre Verwendung in der alleinigen Erfassung zur Beseitigung der Probleme sucht, wird keine explizite Vorgehensweise für eine Gebrauchstauglichkeitsbewertung medizinischer Produkte beschrieben.

#### **Ergonomie-Leitfaden für die Gestaltung von Medizinprodukten**

Der von Hölscher & Laurig entwickelte Ergonomie-Leitfaden für die Gestaltung von Medizinprodukten [186] kann den Autoren nach zur systematischen Berücksichtigung der relevanten ergonomischen Aspekte bei der Gestaltung von Medizinprodukten herangezogen werden. Der Leitfaden soll „Produktnormern“ und Herstellern eine Hilfestellung bei der Berücksichtigung und der Aufnahme ergonomischer Aspekte in Produktnormen sowie bei der Produktgestaltung bieten. Er passt den für die Maschinennormung entwickelten Ergonomie-Leitfaden DIN EN 13861 [95] an die besonderen Anforderungen bei der Gestaltung von Medizinprodukten an. Zur systematischen Evaluation von Risiken geht er von einer Gefährdungsanalyse und einer Risikoeinschätzung aus. Grundlage für den Ergonomie-

Leitfaden für Medizinprodukte ist die Abschätzung relevanter Gefährdungen als Folge einer Vernachlässigung ergonomischer Grundsätze. Durch die Risikobeurteilung wird überprüft, welche der relevanten Gefährdungen spezielle Maßnahmen des Konstrukteurs verlangen, um ein Risiko auszuschließen oder zu reduzieren.

Hierzu liefert er in enger Verbindung zu den gängigen Normen Aspekte zur Risikobeurteilung, -identifizierung, -einschätzung und -bewertung und stellt dem Anwender folgende Werkzeuge zur Verfügung:

- einen Fragenkatalog auf Grundlage der DIN EN ISO 14971 [108] zur Identifizierung von Gefährdungen und zur Ermittlung allgemeiner sicherheitskritischer Bedingungen des Gebrauchs,
- eine Tabelle auf Grundlage der DIN EN 1050 [94] zur Ermittlung von Gefährdungen, die auf einer Vernachlässigung ergonomischer Grundsätze bei der Produktgestaltung begründet sind,
- eine Übersichtsliste mit kurzen Inhaltsangaben ausgewählter Ergonomie-Normen, die dem Konstrukteur im Anschluss an die Gefährdungsanalyse Hilfestellung bei der ergonomischen Gestaltung von Medizinprodukten bieten soll.

Darüber hinaus lässt der Ergonomie-Leitfaden aber spezielle Validierungs- und Entwicklungsempfehlungen für eine Benutzerschnittstelle vermissen.

### **Vergleichende Warentests**

Unter vergleichenden Warentests werden Qualitätsprüfungen bei Gütern verstanden, bei denen kontrollierbare und reproduzierbare gebrauch- und umweltrelevante Produkteigenschaften begutachtet werden [162]. In vergleichenden Warentests „[...] werden vergleichbare, meist konkurrierende Marken anhand identischer Kriterien geprüft und beurteilt“ [356]. Das Ergebnis dieser Tests wird der Öffentlichkeit zum Zwecke der Verbraucheraufklärung bekannt gegeben. Der Terminus „Warentests“ umfasst dabei im allgemeinen Sprachgebrauch sowohl Produkttests als auch Dienstleistungstests. Warentests werden in der Regel von unabhängigen Stiftungen und Institutionen, wie z.B. der „Stiftung Warentest“ durchgeführt und in Zeitschriften wie z.B. „test“ publiziert. Innerhalb dieser Tests werden auch Gebrauchstauglichkeitsaspekte betrachtet [38]. Laut Satzung soll die Stiftung Warentest „um den Stiftungszweck zu verwirklichen, [...] Untersuchungen, in der Regel vergleichender Art, an Waren und Leistungen nach wissenschaftlichen

Methoden und in einem eine sachgerechte Beurteilung gewährleistenden Ausmaß durchführen oder von geeigneten Instituten nach ihren Weisungen durchführen lassen [...]“ [378]. Für die spezielle Prüfung der Gebrauchstauglichkeit existiert hierzu jedoch kein Normblatt [194].

Die typische Durchführung eines Tests der Stiftung Warentest beinhaltet, dass nach marktanalytischen Vorbereitungen und Auswahl der Produkte fünf „erfahrene“ Laien, ausgesucht unter Berücksichtigung des Nutzerprofils und angeleitet durch Fachleute, Erhebungsbögen ausfüllen und entsprechend vorgegebener Prüfpunkte bewerten. Besonders gute oder schlechte Noten werden hierbei begründet [38]. Die Untersuchungsstruktur gliedert sich nach: Funktion (etwa zu 40 bis 50 %), Handhabung (etwa zu 25 bis 30 %), Umwelt und Gesundheit (etwa zu 20 %), Sicherheit (etwa zu 5 bis 10 %), Haltbarkeit (etwa zu 10 bis 20 %) und wird entsprechend dem zu betrachtenden Gegenstand gewichtet. Teilweise werden diese Prüfungen noch durch Expertenbeurteilungen ergänzt [38]. Gebrauchstauglichkeit wird nicht gesondert beachtet, lässt sich aber besonders in den Kriterien „Handhabung“ und „Funktion“ wiederfinden [38].

Die Prüfung erfolgt durch unabhängige Institute. Die Auswertung der Ergebnisse und die Festlegung der Qualitätsurteile nimmt die Stiftung Warentest vor. Testergebnisse werden durch die Stiftung in eine, den Schulnoten entsprechende 5-stufige Bewertungsskala (Qualitätskategorie) eingeordnet [194]. Hierzu werden teilweise Experten (neutrale Fachleute) hinzugezogen. Charakteristikum der Warentests ist die Neutralität durch Unabhängigkeit der durchführenden Testorganisationen und -labore. Diese Unabhängigkeit ist gerade bei kommerziellen Anbietern oft umstritten. Kritische Marktteilnehmer beanstanden, dass die Testergebnisse nicht nachvollziehbar, sachlich falsch oder extern geprägt sind. Für die Stiftung Warentest ist dies jedoch dadurch entkräftet, dass sie in der Öffentlichkeit ein hohes Ansehen genießt, ihre Tests nicht selbst durchführt, die Testmethoden genau dokumentiert und in keiner finanziellen Abhängigkeit zu Dritten steht.

### **Verschiedene Methoden des Usability Engineerings**

Um die Gebrauchstauglichkeit von Geräten zu überprüfen, haben sich zahlreiche Methoden etabliert, welche für spezielle Einsatzgebiete und Produktentwicklungsstufen einsetzbar sind (Kap. 2.1.4 u. 2.4.1). Lediglich für die Betrachtung der Gebrauchstauglichkeit unter dem Aspekt der Sicherheit, besonders im medizinischen Bereich, sind noch nicht viele nennenswerte Methoden untersucht worden.



Die meisten beurteilen lediglich die Gebrauchstauglichkeit, ohne jedoch einen sicherheitskritischen Aspekt und damit verbundene Risiken zu betrachten.

Auch sind viele Kritiker vorhanden, die gerade die qualitativen Methoden der Gebrauchstauglichkeitsmessung anzweifeln. Untersuchungen der Validität der in Kapitel 2.4.1 genannten Methoden stehen hierzu als Gegenargumente [72, 73, 74, 120, 142, 210, 223, 285]. Ebenso sind Kritiker zu finden, welche die Meinung vertreten, dass zwar Gebrauchstauglichkeitsprobleme validierbar sind, jedoch keine Beurteilung oder Bewertung der Gebrauchstauglichkeit von Produkten möglich ist. Richtig ist einerseits, dass bei den meisten Studien, welche den Anspruch erheben, alle oder fast alle Bediendefizite eines Gerätes zu ermitteln, der Aufwand die entstehenden Kosten nicht rechtfertigt, und es somit vorkommen kann, dass mit den getesteten Geräten in der Praxis immer noch Probleme auftreten. Trotzdem gibt andererseits jeder Hinweis auf ein Problem Hilfestellung zur Lösung oder Minderung desselben. Außerdem müssen zur fundierten Bewertung eines Produktes das Gerät nicht bis ins kleinste Detail untersucht und alle möglichen Extremfälle und Situationen inkludiert werden. Spezielle Bewertungsmöglichkeiten für Produkte werden bei den Methoden eher nicht vorgeschlagen, sondern lediglich Bewertungen einzelner Probleme.

Um die Gebrauchstauglichkeit zu ermitteln und zu messen, sind die in der DIN EN 9241-11 [113] definierten Bewertungskriterien Effektivität, Effizienz und Zufriedenheit zu verwenden. Diese globalen Maßstäbe lassen sich noch durch die in Teil 110 beschriebenen Dialogprinzipien konkretisieren (vgl. Kap. 2.4) [114]. Für das entsprechende Produkt müssen dann je nach Ziel der Untersuchungen die entsprechenden Methoden gewählt werden. Eine genaue Abhandlung der Methodenwahl wurde bereits in Kapitel 2.4 vorgestellt.

#### **Methoden des Risikomanagements**

Methoden des Risikomanagements eignen sich, um Risiken zu entdecken und zu bewerten, jedoch lassen sie Ansätze vermissen, welche sich auf Risiken beziehen, die auf Gebrauchstauglichkeitsproblemen beruhen. Methoden wie z.B. HAZOP oder FMEA wurden zwar schon häufig an Designprobleme angepasst, beachten aber nach Hollnagel immer noch nicht ausreichend die menschliche Komponente in einem Arbeitssystem [146, 185]. Da die genaue Abschätzung eines Risikos sehr wichtig ist, sind die Bewertungsmöglichkeiten bei den Methoden des Risikomanagements sehr ausgeprägt. Weiterhin berechnen die Metho-

den des Risikomanagements Risiken vor dem Eintreffen bestimmter Situationen bzw. schätzen sie ein. Bei der Problemstellung in dieser Arbeit werden risikoträchtige Situationen bereits bei einer Gebrauchstauglichkeitsuntersuchung entdeckt. Dennoch können ähnliche Bewertungsschemata herangezogen werden.

### **3.2.2 Diskussion der Methoden und Rückschlüsse für eine Standardprozedur**

Jede der vorgestellten Methoden besitzt spezielle Stärken und Schwächen. Im Krankenhaus wird im Allgemeinen lediglich eine Probestellung durchgeführt. Diese Art der Überprüfung eines Medizinproduktes ist meist weder geplant noch zuverlässig und effektiv [187] und daher wenig aussagekräftig. Während der Probestellung werden in der Regel „kritische Kombinationen von Randbedingungen“ und „arbeitsbedingte Stresssymptome“ nicht beachtet. Daher lässt sich die Gebrauchstauglichkeit eines Produktes laut Hölscher et al. so nur peripher oder gar nicht ermittelt werden [186], wodurch die Anforderungen des MPG und der MPBetreibV (Kap. 2.3) nicht erfüllt werden.

Die meisten der vorgestellten Techniken können einen Beitrag dazu leisten, die Gebrauchstauglichkeit medizinischer Geräte zu bestimmen, jedoch kann keine Methode gänzlich überzeugen. Die Verfahren stellen in der Regel Normabweichungen oder Probleme heraus, um diese in der Entwicklung zu korrigieren, und lassen daher Bewertungsmöglichkeiten vermissen. Das in dieser Arbeit zu entwickelnde Verfahren soll jedoch den Anspruch haben, die Gebrauchstauglichkeit zu bewerten. Immerhin können von den verschiedenen Methoden Ansätze für diese Fragestellung verwendet werden. Die vielversprechendsten Ansätze sind im Methodenrepertoire des Usability Engineerings zu finden, jedoch sind diese nicht speziell auf die Medizin ausgerichtet und lassen keine integrierte und umfassende Analyse sicherheitskritischer Systeme zu [140]. Als sinnvollste Methode für eine Standardprozedur stellt sich entsprechend Kapitel 2.4.2 eine Kombination der Methoden Heuristische Evaluation und Nutzertest heraus.

Für die Risikobewertung der evaluierten Probleme sind Methoden des Risikomanagements geeignet, wenn auch nur in Ansätzen, da sie die Beachtung des Nutzungskontextes und der Mensch-Maschine-Schnittstelle vermissen lassen. Die Ergebnisse der Kommission für Arbeitsschutz (KAN), des Frameworks nach Lowery & Martin, des Ergonomie-Leitfadens für die Gestaltung von Medizinprodukten und der Use Error Checkliste für Evaluationsschwerpunkte können als

Gedankenstützen für eine heuristische Evaluation oder Expertenevaluation verwendet werden. Das Schema nach Hölscher et al. kann herangezogen werden, um Prüfkriterien für einen Gebrauchstauglichkeitstest abzuleiten. Die im DATech Prüfverfahren empfohlene Erhärtungsprüfung für vermutete Normabweichungen ist sinnvoll für eine Standardprozedur zur Ermittlung der Gebrauchstauglichkeit medizinischer Geräte. Die durch eine Expertenevaluation entdeckten, resp. vermuteten Gebrauchstauglichkeitsprobleme können durch die Wahl entsprechender Testszenarien in einem nachfolgenden Gebrauchstauglichkeitstest erhärtet werden. Die Methode PROMEDIKS zeigt, dass die Methoden des Usability Engineering für den medizinischen Bereich anwendbar sind, und gibt hierfür entsprechende Anhaltspunkte. Ein abschließender Methodenvergleich und dessen Relevanz für die in dieser Arbeit zu entwickelnde Methode ist in Tab. 10 dargestellt.

	Beachtung der Gebrauchstauglichkeit	Für Medizinprodukte geeignet	Einfach durchzuführen	Bewertung möglich	Beachtung des Nutzungskontextes
Probestellung					Je nach Durchführung:
FDA "User Error Checklist"					
Framework nach Lowery & Martin					
MECHA					
DATech Prüfhandbuch					
KAN Fragebogen					
Schema nach Hölscher et al.					
Ergonomie-Leitfaden für die Gestaltung von Medizinprodukten					
PROMEDIKS					
Tests wie Stiftung Warentest			unterschiedlich		Eher
Nutzertest					
Methoden des Risikomanagements			unterschiedlich		
Sehr geeignet      nicht geeignet					

**Tab. 10: Relevanz bekannter Prüfmethoden für die in dieser Arbeit zu entwickelnde Methode**

Eine zufriedenstellende Methode, welche die Bewertung der Gebrauchstauglichkeit medizinischer Produkte ermöglicht, gibt es bisher noch nicht. Für deren Ent-

wicklung müssen die jeweiligen Gebrauchstauglichkeitsprobleme weitestgehend erfasst und in einem Folgeschritt bewertet werden. Für das Grundgerüst einer zu entwickelnden Standardprozedur ist die Ermittlung des Nutzungskontextes unabdingbar. Hierdurch werden die relevanten Gebrauchstauglichkeitsanforderungen ermittelt und es können Rückschlüsse auf Gebrauchstauglichkeitsziele gezogen werden. Zur Evaluation medizinischer Geräte werden in dieser Arbeit die vielversprechendsten Evaluationsmethoden Nutzertest und Heuristische Evaluation ausgewählt. Für die Ableitung der Bewertungsmethode werden entsprechende Methoden der Risiko- und Gebrauchstauglichkeitsbewertung hinzugezogen, verglichen und diskutiert. Die folgenden Kapitel behandeln die elementaren Bereiche der Standardprozedur theoretisch, um dann die ausgewählten Methoden und Methodenelemente in einem iterativen Prozess aus Methodenentwicklung und Evaluation zu konkretisieren und zu ergänzen.

### **3.3 Anforderungen an die Methode**

Zu Beginn der Methodenentwicklung müssen Anforderungen an die Methode gestellt werden, mit deren Erfüllung das angestrebte Ziel in vollem Umfang erreicht werden kann. Sie dienen als Unterziele und werden während der Entwicklung der Standardprozedur berücksichtigt und überprüft. Diese Anforderungen gliedern sich in Anforderungen an die Methode selbst und Anforderungen an die Ergebnisse der Methode. Die Anforderungen sind im Einzelnen:

#### **A1. Finden ausreichend vieler Probleme**

Die Standardprozedur soll ausreichend viele Probleme aufdecken, um eine Bewertung des Gerätes zu ermöglichen. Dabei ist es keinesfalls notwendig, dass sie alle Gebrauchstauglichkeitsprobleme aufdeckt, jedoch sollen ausreichend viele und besonders die schwerwiegendsten und sicherheitsrelevanten Probleme aufgedeckt werden, um eine faire und objektive Bewertung zu gewährleisten.

#### **A2. Durchführbarkeit im tatsächlichen Nutzungskontext**

Eine Gebrauchstauglichkeitsprüfung ist aussagekräftig und repräsentativ, wenn sie tatsächliche Nutzer inkludiert, die repräsentative Aufgaben in einer repräsentativen Umgebung erfüllen [5, 105, 240], und wenn sie die von einem Arbeitssystem, Arbeitsplatz oder Arbeitsmittel ausgehenden oder im Umfeld wirksamen Belastungsfaktoren erfasst [358]. Diese Anforderung muss besonders im Bereich

der Medizin gelten, da dieses Arbeitssystem durch besondere Umstände geprägt ist (vgl. Kap 1.2.2.1), welche die Aufgabenerfüllung beeinflussen können.

### **A3. Gebrauchstauglichkeit der Standardprozedur**

Verfahren und Methoden zur Evaluation sollen einfach in der Handhabung sein und auch die Anforderungen der Gebrauchstauglichkeit (Effizienz, Effektivität und Zufriedenheit) erfüllen. Dies gilt insbesondere für Gebrauchstauglichkeitsevaluationsverfahren.

### **A4. Ethische Vertretbarkeit und Vermeidung von Störungen im Arbeitsablauf des Krankenhauses**

Gerade im medizinischen Arbeitssystem ist es wichtig, dass das Verfahren den Arbeitsablauf nicht stört. Durch die Methode entstehende Sicherheitsrisiken für Patient oder Personal sind nicht gerechtfertigt. Weiterhin darf die Methode die Würde von Patienten oder Personal nicht verletzen und muss mit dem Schutz der individuellen Rechte aller Beteiligten vereinbar sein.

### **A5. Angemessenheit bzgl. Kosten-Nutzen-Relation innerhalb der gegebenen Fragestellung**

Die Methode muss für die gegebene Fragestellung angemessen sein. Der durch die Methode entstehende Aufwand an Ressourcen, wie Zeit, Material, Evaluatoren und letztlich an Kosten darf nicht in einem unangebrachten Verhältnis zu dem durch die Methode erhaltenen Nutzen stehen [24, 168].

### **A6. Erwerb einer hohen Akzeptanz**

Die Deutsche Gesellschaft für Evaluation (DeGEval) fordert für Evaluationen eine allgemeine Akzeptanz. Hierzu gehören sowohl ein diplomatisches Vorgehen als auch die Glaubwürdigkeit und Kompetenz des Evaluators [24]. Dementsprechend sollen Evaluationen so geplant und durchgeführt werden, dass sie eine möglichst hohe Akzeptanz der Interessensgruppen erreichen. Dieser Punkt ist vor allem deshalb wichtig, weil die Methode zum einen eine Kooperationsbereitschaft verschiedener Interessenten (Anwender, Betreiber, Hersteller) benötigt, zum anderen ohne diese allgemeine Akzeptanz keinen entsprechenden Nutzen finden würde, und somit die evaluierten Verbesserungspotentiale und Empfehlungen nicht zur Anwendung kommen könnten.

### **A7. Objektivität, Reliabilität und Validität**

Die ISO 25000 fordert Wiederholbarkeit, Reproduzierbarkeit, Unparteilichkeit und Objektivität eines Evaluationsprozesses [199]. Diese Forderungen entsprechen den klassischen psychologischen Testgütekriterien der Reliabilität und Objektivität [168]. Weiterhin sollte Validität gegeben sein, welche nach Harloff bei einer Gebrauchstauglichkeitsevaluation durch eine Ermittlung der in der DIN EN 9241-11 definierten Gebrauchstauglichkeitskriterien Effektivität, Effizienz und Zufriedenheit gewährleistet ist [168].

### **A8. Nützlichkeit**

Die Methode soll einen Beitrag zur Optimierung des medizinischen Arbeitssystems liefern. Oft werden aufwendig und professionell erhobene Ergebnisse nicht entsprechend genutzt. Gemäß der Forderung der Deutschen Gesellschaft für Evaluation soll die Methode den am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen nutzen. Hierzu müssen deren Interessen geklärt und entsprechend bei der Evaluation berücksichtigt werden. Planung, Durchführung und Berichterstattung soll die Beteiligten dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen [24]. Die Methode soll, wie von Schmidtke gefordert, einem praktischen Bedürfnis entsprechen, wobei die Erfassung durch kein anderes Instrumentarium gleich gut möglich ist [358].

### **A9. Fairness und Neutralität (Unparteiische Durchführung und Berichterstattung)**

Gerade unter dem Aspekt, dass bei einer Veröffentlichung von Negativergebnissen entscheidend in den Markt eingegriffen werden kann, ist der Gesichtspunkt „Fairness und Neutralität“ von besonderer Bedeutung. Die Methode soll unparteiisch in ihrer Durchführung und Berichterstattung und nicht durch Lobbyisten geprägt sein.

### **A10. Offenlegung der Ergebnisse und Methoden (Transparenz)**

Die Evaluationsergebnisse sollen allen Beteiligten, Betroffenen und Interessenten soweit wie möglich zugänglich gemacht werden [24]. Hierbei ist wichtig, dass eine Transparenz über Methoden und Ergebnisse gegeben ist, damit diese nachvollziehbar und verständlich sind. Die Methode soll auf einer theoretischen Modellvorstellung aufbauen, welche dem Anwender, Betreiber und Hersteller des Gerätes eine anwendungsorientierte Ergebnisinterpretation möglich macht [206].

**A11. Berücksichtigung des Standes der Forschung und Technik**

Die Methode soll dem aktuellen Stand der wissenschaftlichen Forschung und der Technik entsprechen. Anerkannte Standards und Methoden sollen berücksichtigt und einbezogen werden.

**A12. Einsetzbarkeit für eine Vielzahl medizinischer Geräte**

Die Methode sollte für eine Vielzahl medizinischer Geräte einsetzbar sein. Somit sollten auch passive Geräte, Geräte für den Heimbereich etc. einbezogen werden. Daher sollte es keine Einschränkung bzgl. Interaktionsgrad, Anwendungsort oder Nutzer geben.

**A13. Abdeckung des versehentlichen Missbrauchs**

Neben dem normalen Gebrauch der Geräte soll die Standardprozedur auch den vorhersehbaren, versehentlichen Missbrauch entsprechend der DIN EN IEC 60601-1-6 berücksichtigen.

**A14. Berücksichtigung der Anforderungen an ein sicherheitskritisches System**

Das medizinische Arbeitssystem ist von Sicherheitsrisiken geprägt. Aus diesem Grund soll die Methode ein besonderes Augenmerk auf die Sicherheit bei der Bewertung der Gebrauchstauglichkeit medizinischer Geräte legen.

**A15. Bewertung der Gebrauchstauglichkeit**

Als primäre Anforderung gilt für die Methode die Bewertung der Gebrauchstauglichkeit medizinischer Geräte. Hierfür sollen die in den Normen DIN EN 9241-11 und DIN EN IEC 60601-1-6 (vgl. Kap. 2.4) definierten Maße der Gebrauchstauglichkeit verwendet werden.

## 4. Entwicklung der Evaluationsmethode

In diesem Kapitel wird eine Evaluationsmethode für Gebrauchstauglichkeitsprobleme in der Medizin entwickelt und evaluiert. Danach wird in Kapitel 5 und 6 eine Bewertungsmethode erarbeitet. Diese getrennte Vorgehensweise dient dazu, den Betrachtungsfokus schwerpunktmäßig auf diese beiden elementaren Bestandteile der Standardprozedur richten zu können.

Die Natur der Gebrauchstauglichkeitsevaluation benötigt keine grundsätzliche Adaption an das medizinische Arbeitssystem. Evaluationsmethoden müssen lediglich an die speziellen Anforderungen dieses Systems angepasst werden. Die Art und Weise dieser Anpassung wird im Folgenden beschrieben und evaluiert.

### 4.1 Material und Methode

In diesem Kapitel werden die elementaren Teile der Standardprozedur beschrieben und diskutiert. Spezielle Anforderungen des medizinischen Nutzungskontextes werden beachtet und in die Entwicklung der Standardprozedur integriert. Als grundsätzliches Rahmenwerk werden die Elemente des Usability Engineering Lifecycles (vgl. Kap. 2.1.4) verwendet und an die Anforderungen des medizinischen Arbeitssystems angepasst. Als Grundlage hierfür dienen die in Kapitel 1 und 2 ermittelten theoretischen Grundlagen, Überlegungen und Diskussionen sowie die in Kapitel 3 aufgestellten Anforderungen an die Standardprozedur, welche kritisch und methodisch auf ihre Relevanz für das medizinische Arbeitssystem überprüft und weiterentwickelt werden. Die speziellen Fragestellungen hierbei sind:

- Wie kann der Benutzungskontext entwickelt werden - welche speziellen Anforderungen ergeben sich aus dem medizinischen Arbeitssystem?
- Wie kann die Gebrauchstauglichkeit gemessen werden - welche Methoden sind geeignet?
- Welche Stichprobe soll verwendet werden?
- Welche Anforderungen werden an den Evaluator gestellt?
- Welche Testszenarien und Aufgaben sollen verwendet werden?
- Wie ist die Qualität des Tests?
- Wie ist der Test aufgebaut – in welcher Testumgebung findet der Test statt?
- Wie werden die Ergebnisse dokumentiert?



Zu jeder Fragestellung werden entsprechende Thesen aufgestellt, welche in der anschließenden Evaluationsphase in einem iterativen Prozess validiert werden, um die Methode und deren Teilelemente an das medizinische Arbeitssystem anzupassen. Im Anhang D werden diese Thesen zusammengefasst dargestellt.

#### 4.1.1 Anforderungsanalyse für medizinische Produkte

*„Der Glaube, dass die eigene Sicht der Wirklichkeit die Wirklichkeit schlechthin bedeutet, ist eine gefährliche Wahnidee. Die sogenannte Wirklichkeit ist das Ergebnis von Kommunikation.“ P. Watzlawick*

Um die Gebrauchstauglichkeit in der Medizin zu untersuchen, ist es wichtig, die Größen zu kennen, welche die Gebrauchstauglichkeit eines Produkts beeinflussen können. Diese werden unter dem Begriff Benutzungskontext [97] oder auch Nutzungskontext [113] (engl. context of use) zusammengefasst. Ein Produkt kann nicht ohne Berücksichtigung des Nutzungskontextes gebrauchstauglich gestaltet bzw. hinsichtlich seiner Gebrauchstauglichkeit bewertet werden [248]. Zentraler Bestandteil des Nutzungskontextes sind der **Benutzer** und die **Arbeitsaufgabe**, welche das Leitziel für die Interaktion mit dem Produkt bilden. Beeinflussende Faktoren sind **Arbeitsmittel**, welche zur Zielerreichung verwendet werden, sowie die organisatorische und physische **Arbeitsumgebung**, in der das Produkt benutzt wird (Abb. 35).



Abb. 35: Zentrale Bestandteile des medizinischen Nutzungskontextes

Die Nutzungskontextanalyse und -beschreibung sind für die Gebrauchstauglichkeitsbetrachtung unerlässlich [33, 97, 105, 113, 202, 248, 346, 426, 430], jedoch hängt ihr Detaillierungsniveau von der jeweiligen Situation und Umgebung ab, in der das zu untersuchende Produkt verwendet wird [105, 113, 223, 334]. Für eine Gebrauchstauglichkeitsbewertung muss eine Nutzungskontextanalyse nur so weit durchgeführt werden, dass die beeinflussenden Faktoren für die Gebrauchstauglichkeit eines Produktes erfasst werden. Generell sollte die Beschreibung des Nutzungskontextes laut der DIN EN ISO 13407 den Arbeitsbereich ausreichend detailliert festlegen, aus geeigneten Quellen abgeleitet und dokumentiert sowie durch den Benutzer bestätigt sein [105]. Als geeignete Methoden und Techniken sind nach der DIN EN 6385 der Einsatz von Instrumenten zur Bewertung der Arbeitsbedingungen, Beobachtungen vor Ort und Befragungen anzusehen [109]. Eine weitere Methode ist die Kontextsitzung<sup>34</sup>, bei der alle Informationen über den Nutzungskontext durch ein interdisziplinäres Team (Produktmanager, Nutzer, Entwickler, Ergonomieexperten) gesammelt, diskutiert und dokumentiert werden [248, 387]. Sollten bereits ähnliche Systeme wie das betrachtete bestehen, so sollten auch Informationen über ergonomische Schwachstellen dieser Systeme erhoben werden. Nach Sammlung und Analyse dieser Informationen sollte ein Katalog mit einer Beschreibung der relevanten Benutzermerkmale, der Arbeitsaufgaben und der Umgebung erstellt werden. Dieser legt fest, welche Aspekte einen wichtigen Einfluss auf die Systemgestaltung haben. Hieraus werden die Anforderungen an das Produkt und die Gebrauchstauglichkeitsziele bestimmt, welche im späteren Verlauf in Evaluationskriterien überführt werden.

<b>Thesen:</b>	
T.1	Der medizinische Nutzungskontext ist durch <b>Benutzer, Arbeitsaufgabe, Arbeitsmittel</b> und <b>Arbeitsumgebung</b> bestimmt.
T.2	Für die Überprüfung der Gebrauchstauglichkeit medizinischer Geräte ist die Erfassung der beeinflussenden Faktoren des Nutzungskontextes unerlässlich.

Im Folgenden werden die einzelnen Elemente des Nutzungskontextes, deren Relevanz für die Gebrauchstauglichkeitsevaluation und deren Analyse genauer beschrieben.

<sup>34</sup> Im Rahmen des Projekts MUSiC wurde ein Leitfaden zur Nutzungskontextanalyse entwickelt [387]

### 4.1.1.1 Nutzerprofil

Jedes Gerät muss die Anforderungen der Anwender bzgl. Zweckdienlichkeit und Zielstellung erfüllen (vgl. Kap. 2.1.3). Bei der Bestimmung des Nutzerprofils muss herausgefunden werden, wer die Anwender sind, welche Funktionalitäten sie erwarten, welche Vorkenntnisse sie mitbringen, was sie zu lernen bereit sind und welche Charakteristiken sie haben. Um es mit dem schon 1971 verwendeten einfachen Slogan nach Hansen zu nennen: „Know the user!“ [167]. Zu diesem Zweck ist ein Nutzerprofil, auch Anwenderprofil oder im Englischen „User Model“ genannt, geeignet. Es stellt eine Sammlung von Informationen über den Nutzer dar, die nach relevanten Merkmalen der Benutzerschnittstelle zusammengestellt ist. Es können sowohl allgemeine Informationen [97, 105, 167, 173, 223, 259, 368] als auch Informationen über spezielle Benutzer enthalten sein. Benutzer werden hierbei in direkte Benutzer, also Nutzer, die mit dem Gerät arbeiten, und indirekte Benutzer, die das Gerät installieren, warten oder administrieren, unterteilt. Letztere Gruppe wird in der Praxis bei Gebrauchstauglichkeitsbetrachtungen fälschlicherweise gar nicht oder nicht ausreichend berücksichtigt.

Für ein Benutzerprofil wird die Zielgruppe hinsichtlich relevanter Merkmale für das betreffende Gerät untersucht, wie z.B. Alter, Geschlecht, körperliche Fähigkeiten, Ausbildung, Erfahrung, Vorlieben, Motivation, Computervertrautheit, Intensivität und Frequenz der Anwendung, Kulturkreis, Vokabular, Sprache, Kenntnisstand der verwendeten Technologie, intellektuelle Fähigkeiten, Einstellungen zum Produkt, zur Aufgabe oder zur Position im Arbeitssystem etc. [97, 105, 160, 173, 223, 259, 368, 414]. Sollten hierbei Nutzergruppen erkennbar sein, wie z.B. Experten und Laien oder häufige bzw. seltene Nutzer, so sollten innerhalb der Gebrauchstauglichkeitsevaluation für die entsprechende Gruppe verschiedene Testmaßstäbe aufgestellt, und wenn nötig, auch verschiedene Tests durchgeführt werden. Hilfreich kann hierbei eine Nutzer/Aufgaben-Matrix (Tab. 11) [160] sein, welche die Aufgaben-Nutzerbeziehungen visualisiert.

	Arzt		Pflegepersonal	Patient	Wartungspersonal
	erfahren	unerfahren			
Aufgabe 1	x		x		
Aufgabe 2	x	x	x	x	
Aufgabe 3			x	x	
Aufgabe 4					x

Tab. 11: Beispiel für eine Nutzer/Aufgaben Matrix

Weiterhin sollte innerhalb eines Tests beachtet werden, dass medizinische Geräte von allen Nutzertypen des Arbeitssystems bedient werden müssen. Somit sollten auch die grenzwertigsten Eigenschaften der potentiellen Nutzer als Maß für einen Gebrauchstauglichkeitstest gelten, wie z.B. der niedrigstmögliche Ausbildungsstand, die geringsten Vorkenntnisse oder das höchste Alter. Mögliche Methoden zur Entwicklung des Nutzerprofils sind Literaturrecherche, Beobachtung, Nutzerinterview, Fragebogenbefragung, Hersteller- und Vertreiberinformationen oder Experteninterviews.

<b>Thesen:</b>	
T.3	Methoden zur Entwicklung des Nutzerprofils sind Literaturrecherche, Beobachtung, Nutzerinterview, Fragebogenbefragung, Hersteller- und Vertreiberinformationen oder Experteninterviews.
T.4	Eine Ermittlung und Unterteilung der Nutzergruppen ist notwendig.
T.5	Unterschiedliche Nutzer medizinischer Geräte benutzen diese anders und unterschiedliche Ergebnisse werden evaluiert.

#### 4.1.1.2 Funktions- und Aufgabenanalyse

*„Die Kunden wollen keine Bohrer sondern Löcher“ C. Rudolf*

*„Schau was deine Nutzer machen, aber frage sie nie, was sie wollen“ J. Raskin*

Um ein Gerät hinsichtlich seiner Gebrauchstauglichkeit zu untersuchen, ist es nicht ausreichend, lediglich die Anwender zu kennen. Auch die Aufgaben der Nutzer müssen evaluiert werden. Hierzu gehört ein Verständnis darüber, welche Ziele die Nutzer mit dem Produkt erreichen wollen, welche Aufgaben und Arbeitsschritte die Nutzer erledigen, wie sie diese erledigen, oder wünschen diese zu erledigen, und welche Probleme sie bei der Erledigung haben.

Die Aufgabenanalyse (engl. Task Analysis) ist eine der wichtigsten Methoden der Mensch-Maschine-Kommunikation [82], welche dem Verständnis und der Beschreibung der Benutzerziele, der Benutzeraufgabe und des Arbeitsablaufs mit einem Produkt in einem Anwendungsbereich dient [15, 67, 160, 207, 284, 261, 313]. Bei der Aufgabenanalyse werden die sequentiellen, manuellen und intellektuellen Aktivitäten der Anwender systematisch gesammelt [97]. Ziel ist es, Arbeitsschritte zu verstehen, relevante Aufgaben zu identifizieren [261, 302] und im Falle einer Produktbewertung in einem Nutzertest zu evaluieren. Deshalb wird sie

auch von Diaper & Stanton als die Wurzel einer Evaluation in der Mensch-Maschine-Kommunikation angesehen [84].

Die Aufgabenanalyse ist nicht eine einzige Methode, sondern besteht aus unterschiedlichen, oft auch variierenden systematischen Methoden, welche für die speziellen Evaluationsziele angepasst werden. Beispiele hierfür sind: allgemeine Aufgabenanalyse [97, 217], Zeit- und Bewegungsstudien [97, 207], hierarchische Aufgabenanalyse [69, 207, 217], kontextuelle Aufgabenanalyse [261], Workflow Analysen [160], kognitive Modelle wie z.B. GOMS (Gesamtziel, Operatoren, Methoden, Selektionsregeln) [52, 69, 323] oder die weiter abstrakte kognitive Aufgabenanalyse<sup>35</sup> [64, 69, 97, 122, 313]. Als Datenerhebungstechniken der einzelnen Methoden können Befragungen von Nutzern und Herstellern, Beobachtungen des Anwenders, Brainstormingtechniken, Fokusgruppen und/oder Einschleusungen von Beobachtern in die Arbeitsumgebung angewendet werden. Wichtige Elemente der Aufgabenanalyse sind im Folgenden dargestellt:

- Ziele der Nutzer
- Häufigkeit der Verwendung
- Zweck der Verwendung
- Typische Aufgaben
- Wichtige und wesentliche Aufgaben
- Aufgaben im Notfall
- Arbeitsabläufe
- Irrelevante Aufgaben
- Unnötig komplexe Aufgaben
- Optimierungsbedarf
- Zeiten für einzelne Aufgaben
- Wahlfreiheit (Hat der Benutzer die Wahl, das Produkt zu verwenden?)
- Sicherheitskritische Aufgaben

Bei der Datenerhebung einer Aufgabenanalyse sollte beachtet werden, dass Nutzer oft Schwierigkeiten haben, ihre Bedürfnisse, Ziele und Aufgaben einzuschätzen und zu verbalisieren. Daher sollten bei einer Aufgabenanalyse Interviews und Feldbeobachtungen kombiniert werden.

Häufig wird eine Aufgabenanalyse bereits in frühen Entwicklungsstadien eingesetzt, um aufgabenbezogene Anforderungen abzuleiten und in den Entwicklungsprozess zu integrieren. Bei einer reinen Bewertung eines Produktes ist dies nicht notwendig. Es reicht aus, die Ziele der Nutzer, die Arbeitsabläufe und Benutzeraufgaben zu identifizieren, zu verstehen, zu evaluieren und ggf. zu kategorisieren. Somit ist in diesem Kontext auch Carrolls an sich richtiger Einwand unbegründet, dass es eine Reihe an potentiellen und infiniten Möglichkeiten gibt, mit denen unterschiedliche Menschen mehr oder weniger erfolgreich und effizient Aufgaben

---

<sup>35</sup> Die kognitive Aufgabenanalyse ist fokussiert auf die kognitiven Prozesse des Anwenders, wie zum Beispiel seine Gedankenmodelle zum System- oder Gerätebetrieb.

erfüllen [53], da bei diesem Ansatz zur Bewertung eines Produktes nur die möglichen Wege aufgezeigt werden, die dem Nutzer durch das Produkt angeboten werden. Es muss überprüft werden, ob diese Wege mit den Wünschen und Vorstellungen der Nutzer übereinstimmen.

Zu diesem Zweck ist besonders die dem Top-Down-Prinzip folgende hierarchische Aufgabenanalyse (HTA = Hierarchical Task Analysis) geeignet [4, 339]. Die Grundidee lehnt sich an die Eigenschaft des Menschen an, ein Problem in kleinere Teilprobleme zu zerlegen [4, 313, 339] und besteht darin, dass die Aufgaben (Tasks) in Unteraufgaben (Sub-Tasks) und Aktionen (Actions) unterteilt und grafisch dargestellt werden [69, 207, 217]. Zusätzlich kann in diese Baumstruktur noch die Zielebene eingefügt werden. Sie umfasst die Ziele, die der Anwender mit den zu erledigenden Aufgaben zu erreichen versucht. Diese werden in Teilziele (Aufgaben, welche zur Zielerreichung benötigt werden) und schließlich in Aktionen zerlegt (vgl. Kap. 1.2.1.1) [46, 113]. Die Teilziele spezifizieren Komponenten eines übergeordneten Zieles sowie die Kriterien der Zielerreichung. Aktionen sind atomare, nicht mehr weiter unterteilbare Nutzeraufgaben, welche keine Problemlösungs- oder Kontrollfunktion besitzen (z.B. Knopf drücken, Wert eingeben) [46, 218]. Auf diese Weise können auch komplexe Nutzeraufgaben zerlegt und anschaulich präsentiert werden.

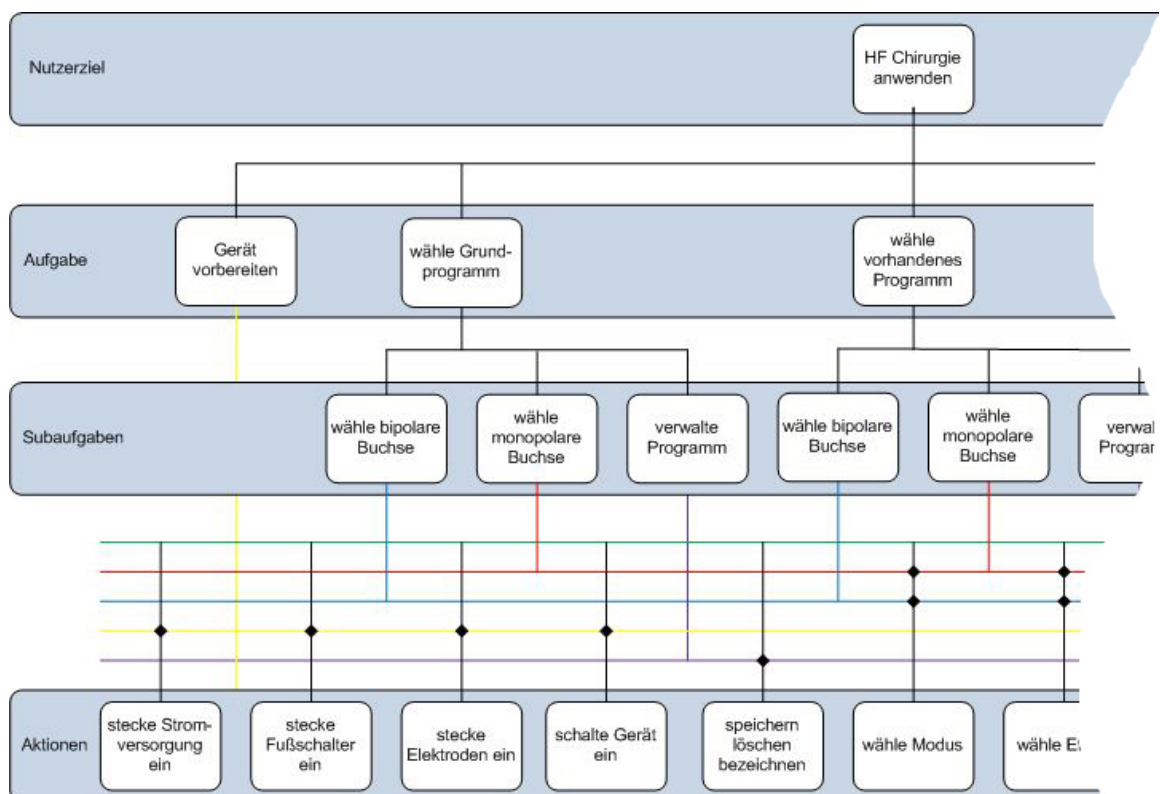


Abb. 36: Beispielausschnitt einer HTA für ein Hochfrequenzchirurgiegerät

Die HTA erlaubt eine Betrachtung der einzelnen Teile von Aufgaben, ohne hierbei den Kontext der ganzen Aufgabe und des Ziels zu verlieren, jedoch wird auf die eigentliche Aufgabenerfüllung nur begrenzt eingegangen. Deswegen kann die HTA auch mit weiteren Analysemethoden ergänzt werden, beispielsweise durch einen ConcurTaskTree (CTT), eine von Paternò aus der LOTUS Notation entwickelte Methode zur Notation von Aufgabenmodellen [302]. Die Eigenschaften dieser Methode umfassen einen Fokus auf die Aktivitäten, eine hierarchische Struktur, eine grafische Syntax und ein reichhaltiges Set an zeitlichen Operatoren, Aufgabenanweisungen, sowie Objekten und Aufgabenattributen [303].

In manchen komplexen Systemen ist es nicht oder nur mit sehr hohem Aufwand möglich, eine HTA bis in das letzte Detaillierungsniveau durchzuführen. Generell sollte eine HTA auch nur soweit gehen, bis die gesammelten Informationen für das Ziel der Analyse ausreichen und ein hinreichendes Verständnis der Nutzeraufgaben erreicht wird. Hierzu zählt mindestens eine vollständige Durchdringung der Primär- und der häufigsten Sekundäraufgaben. Wird eine HTA für die Bewertung der Gebrauchstauglichkeit eines Medizinproduktes verwendet, ist eine Kategorisierung der Ziele, Aufgaben und Aktionen sinnvoll. Eine Möglichkeit ist die in Kapitel 1.2.2.2 beschriebene Einteilung in Primär- und Sekundäraufgaben.

<b>Thesen:</b>	
T.6	Als Methoden für ein Aufgabenprofil kommen Befragungen (Interviews, Fragebögen) von Nutzern und Herstellern, Beobachtungen des Anwenders, Brainstorming-Techniken, Fokusgruppen und/oder Kontextbegehungen in Betracht.
T.7	Mit der Aufgabenanalyse können die mit dem System durchgeführten Ziele, Aufgaben und Aktionen der Nutzer ermittelt und kategorisiert werden. Sie liefert somit elementare Informationen für die Gebrauchstauglichkeitsprüfung.
T.8	Aufgaben sollten für die Evaluation in Primär- und Sekundäraufgaben unterteilt werden.

#### 4.1.1.3 Analyse der Arbeitsumgebung

Die Arbeitsumgebung, in der ein Produkt verwendet wird, ist von besonderer Bedeutung. Grundsätzlich gilt sowohl für eine Gebrauchstauglichkeitsprüfung als auch für eine Produktentwicklung, dass der Anwender mindestens einmal in seiner realen Arbeitsumgebung observiert wird, um ein Gefühl für die Verwendung des Gerätes in seiner realen Umgebung zu bekommen.

Die Arbeitsumgebung beschreibt die Faktoren, welche auf den arbeitenden Menschen leistungsbeeinflussend wirken. Sobald auch nur ein Faktor der Arbeitsumgebung verändert wird, können andere Anforderungen an das Produkt gelten. Im Allgemeinen wird die Arbeitsumgebung in eine physische, organisatorische und soziale Arbeitsumgebung unterteilt. Die physische Arbeitsumgebung umfasst die Umgebungseinflüsse (z.B. Beleuchtung, Klima, Akustik, mechanische Schwingungen, Schmutz,) und den daraus resultierenden Gefährdungscharakter. Zeitliche Arbeitsorganisation, Arbeitsabläufe, Verantwortung und gesetzliche Richtlinien werden durch die organisatorische und soziale Arbeitsumgebung erfasst. Zusätzlich schließt die Arbeitsumgebung Hardware, Software und die verwendeten Materialien mit ein [105, 160]. Auch die kulturelle Umgebung kann in manchen Fällen eine Rolle spielen. Die Beschreibung der Arbeitsumgebung enthält eine vollständige Ermittlung und Darstellung des physischen und sozialen Umfeldes sowie eine Beschreibung der Leistungseigenschaften und Merkmale der Arbeitsmittel, welche die Gebrauchstauglichkeit eines Produktes beeinflussen. Ergonomische Anforderungen für die Arbeitsumgebung, die dem Benutzer bequeme, sichere und produktive Arbeitsbedingungen bieten, werden in der internationalen Norm DIN EN 9241-6 festgelegt [116]. Eine spezielle Norm für medizinische Geräte gibt es bisher nicht. Lediglich die DIN EN IEC 60601-1-6 (vgl. Kap. 2.3.2) verweist auf die folgenden Faktoren, welche in diesem Zusammenhang berücksichtigt werden sollen [97]:

- räumlicher Zusammenhang
- sozialer Zusammenhang
- technologischer Zusammenhang
- hygienischer Zusammenhang
- physischer Zusammenhang
- Aktivitätszusammenhang

Für medizinische Geräte gilt, dass die Arbeitsumgebung und somit auch die Anforderungen an diese Produkte sehr unterschiedlich sind. Viele von ihnen finden nicht nur im Krankenhaus ihre Anwendung, sie kommen im häuslichen Bereich oder universell zum Einsatz oder können sogar in einer speziellen Umgebung wie z.B. auf einem Schiff oder gar im Weltall ihre Aufgabe erfüllen. Der Nutzer kann sich bei der Bedienung in völlig unterschiedlicher psychischer und physischer Verfassung befinden [334]. In den unterschiedlichen Anwendungssituationen und Anwendungsumgebungen ändert sich auch der Nutzungskontext. Bei den Arbeitsmitteln sollten neben der Hard- und der Software auch die Materialien und die Kompatibilität der Geräte betrachtet werden. Wichtig hierbei ist der



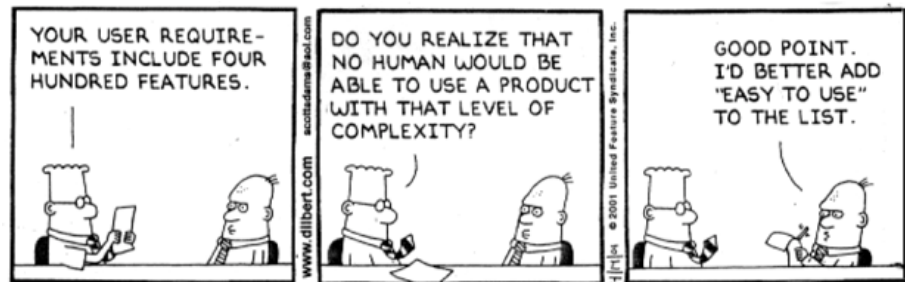
Systemgedanke. Die meisten Geräte sind Teil eines Systems oder werden in einem System verwendet. Hierbei muss der Nutzungskontext auch die Eigenschaften und die Kompatibilität der anderen relevanten Elemente des Systems berücksichtigen. Für diese Systemintegration legt der Normentwurf DIN IEC 80001 einen Prozess fest, der „Maßnahmen und besonders das Risikomanagement umfasst, welche von der verantwortlichen Organisation und dem Hersteller gefordert werden, wenn Medizinprodukte in ein Netzwerk eingebunden werden oder derartige Einbindungen verändert werden“ [129]. Hilfreich bei der Evaluation der Arbeitsumgebung können die von Basyat et al. näher beschriebenen Umwelt- (engl. Environment models), Plattform-, System- und Gerätemodelle (engl. Device models) sein [15]. Tab. 12 fasst die wichtigsten Komponenten des Nutzungskontextes in der Medizin zusammen.

Arbeitsmittel	Nutzer	Aufgabe
<b>Allgemein:</b> <ul style="list-style-type: none"> <li>- System oder Produktname und Version</li> <li>- System oder Produktbeschreibung</li> <li>- Anwendungsbereich</li> <li>- Hauptfunktionen</li> <li>- Zielmarkt</li> <li>- Interessensvertreter</li> </ul> <b>Spezifikation</b> <ul style="list-style-type: none"> <li>- Materialien</li> <li>- Dienstleistungen</li> </ul>	<b>Benutzertyp</b> <ul style="list-style-type: none"> <li>- Primär- und Sekundärnutzer</li> <li>- Benutzerrolle</li> </ul> <b>Fertigkeiten und Wissen</b> <ul style="list-style-type: none"> <li>- Laie und Experte</li> <li>- Einweisung / Training</li> <li>- Qualifikation</li> <li>- Sprachkenntnisse</li> </ul> <b>Persönliche Merkmale</b> <ul style="list-style-type: none"> <li>- Alter, Geschlecht</li> <li>- Physische Fähigkeiten</li> <li>- Grenzen und Behinderungen</li> <li>- Einstellung u. Motivation</li> </ul>	<ul style="list-style-type: none"> <li>- Nutzerziele</li> <li>- Primär- o. Sekundäraufgabe</li> <li>- Aufgaben zur Zielerreichung, Unteraufgaben und Aktionen</li> <li>- Aufgabenhäufigkeit</li> <li>- Aufgabendauer</li> <li>- Aufgabenziel</li> <li>- Physische und mentale Anforderungen</li> <li>- Aufgabenergebnisse</li> <li>- Aufgabenabhängigkeit</li> <li>- Auswirkungen von Fehlern</li> <li>- Sicherheitskritische Erfordernisse</li> </ul>
<b>Arbeitsumgebung</b>		
Technische Umgebung	Physische Umgebung	Organisatorische Umgebung
<ul style="list-style-type: none"> <li>- Hardware</li> <li>- Software</li> <li>- System</li> <li>- Kompatibilität</li> <li>- Netzwerk</li> <li>- Weiteres Equipment</li> <li>- Referenzmaterial</li> </ul>	<b>Arbeitsplatzbedingungen</b> <ul style="list-style-type: none"> <li>- Atmosphärische Bedingungen</li> <li>- Akustische Bedingungen</li> <li>- Thermische Bedingungen</li> <li>- Wahrnehmungsbedingungen</li> <li>- Umgebungsstabilität</li> </ul> <b>Arbeitsplatzgestaltung</b> <ul style="list-style-type: none"> <li>- Größe und Einrichtung</li> <li>- Körperhaltung</li> <li>- Arbeitsplatz</li> </ul> <b>Arbeitsplatzsicherheit</b> <ul style="list-style-type: none"> <li>- Gesundheitsgefährdung</li> <li>- Patientengefährdung</li> <li>- Schutzkleidung und -vorrichtungen</li> </ul>	<b>Struktur</b> <ul style="list-style-type: none"> <li>- Arbeitsstunden</li> <li>- Gruppenarbeit</li> <li>- Funktion</li> <li>- Arbeitspraxis</li> <li>- Unterbrechungen</li> <li>- Kommunikationsstruktur</li> </ul> <b>Einstellung und Kultur</b> <ul style="list-style-type: none"> <li>- Vorschriften</li> <li>- Organisatorische Ziele</li> <li>- Geschäftsbeziehungen</li> </ul> <b>Arbeitsgestaltung</b> <ul style="list-style-type: none"> <li>- Leistung</li> <li>- Ergebnismeldung</li> <li>- Arbeitstempo</li> <li>- Selbständigkeit</li> <li>- Entscheidungsfreiheit</li> <li>- Stress, Zeit- oder Sozialdruck</li> </ul>

Tab. 12: Zusammenfassung einiger Komponenten des Nutzungskontextes in der Medizin [113, 250]

Thesen:	
T.9	Die Analyse der Arbeitsumgebung und der Arbeitsmittel liefert elementare Informationen für den Aufbau der Gebrauchstauglichkeitsprüfung.
T.10	Die Umgebungsanalyse liefert Informationen darüber, welche Faktoren als leistungsbeeinflussend bei der Gebrauchstauglichkeitsevaluation berücksichtigt werden müssen.

#### 4.1.2 Prüfkriterien – Messung der Gebrauchstauglichkeit



[Scott Adams]

Aus den Daten des im vorherigen Kapitel beschriebenen Nutzungskontextes lassen sich Anforderungen an das Produkt und Gebrauchstauglichkeitsziele festlegen. Auf deren Grundlage können explizite Evaluationskriterien definiert werden (Abb. 37), um sie in einer Evaluation hinsichtlich Erfüllung bzw. Erfüllungsgrad zu überprüfen [248]. Hierzu müssen die Gebrauchstauglichkeitsziele operationalisiert und den entsprechenden Kriterien zugewiesen werden (vgl. Kap. 2.4). Die aufgestellten Ziele können hierbei sowohl allgemein für das Gerät (z.B. intuitiv zu bedienen, leicht zu erlernen) als auch für spezielle Teilbereiche oder Aufgaben (z.B. ein Medikament schnell und einfach zu verabreichen) gelten.

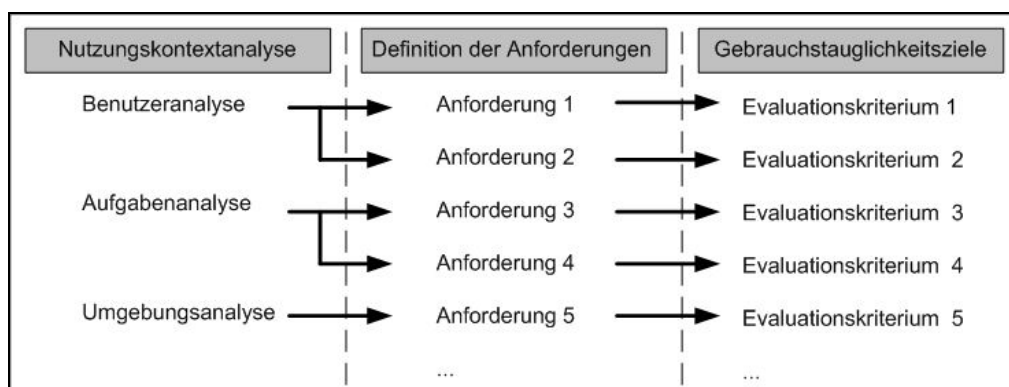


Abb. 37: Vom Nutzungskontext zu den Evaluationskriterien [Adaptiert nach 248]

Zur Messung werden den Gebrauchstauglichkeitsmaßen Effektivität, Effizienz und Zufriedenheit quantitativ messbare Aussagen zugewiesen (Tab. 13). Mit diesen Prüfkriterien kann die Gebrauchstauglichkeit eines medizinischen Produktes

gemessen werden. Dieses Kapitel handelt von der praktischen Messung der Gebrauchstauglichkeit und der hierfür notwendigen Methodenselektion für die Standardprozedur.

Ziel der Gebrauchstauglichkeit	Maße der Effektivität	Maße der Effizienz	Maße der Zufriedenheit
Allgemeine Gebrauchstauglichkeit	<ul style="list-style-type: none"> <li>- Grad der Zielerreichung</li> <li>- Prozentsatz der Benutzer, die die Aufgabe erfolgreich abschließen</li> <li>- Durchschnittliche Genauigkeit der abgeschlossenen Aufgabe</li> </ul>	<ul style="list-style-type: none"> <li>- Zeit für Erledigung einer Aufgabe</li> <li>- Abgeschlossene Aufgaben pro Zeiteinheit</li> <li>- Monetäre Kosten der Aufgabenerledigung</li> </ul>	<ul style="list-style-type: none"> <li>- Einstufung der Zufriedenstellung durch Fragebögen</li> <li>- Häufigkeit von Beschwerden</li> </ul>
Angemessen für erfahrene Benutzer	<ul style="list-style-type: none"> <li>- Anzahl ausgeführter Aufgaben mit hoher Schwierigkeit</li> <li>- Prozentsatz relevanter Funktionen, die genutzt werden</li> </ul>	<ul style="list-style-type: none"> <li>- Effizienz im Verhältnis zu einem Experten-Nutzer</li> </ul>	<ul style="list-style-type: none"> <li>- Einstufungs-Skala für Zufriedenstellung mit Produktmerkmalen für hohen Anspruch</li> </ul>
Angemessen für unerfahrene Benutzer	<ul style="list-style-type: none"> <li>- Prozentsatz erfolgreich abgeschlossener Primäraufgaben</li> </ul>	<ul style="list-style-type: none"> <li>- Erforderliche Zeit für den ersten Versuch</li> <li>- Relative Effizienz beim ersten Versuch</li> </ul>	<ul style="list-style-type: none"> <li>- Einstufungs-Skala für Zufriedenstellung</li> </ul>
Fehlertoleranz	<ul style="list-style-type: none"> <li>- Prozentsatz der Fehler, die vom System erkannt oder berichtet wurden</li> <li>- Anzahl der vom System tolerierten Fehler des Benutzers</li> </ul>	<ul style="list-style-type: none"> <li>- Zeitaufwand zur Fehlerkorrektur</li> </ul>	<ul style="list-style-type: none"> <li>- Einstufungs-Skala für subjektiv erlebte Fehlerbehandlung</li> </ul>

**Tab. 13: Beispiele für Maße der Gebrauchstauglichkeit [113]**

Aus Kapitel 2.4, insbesondere aus den mit Hilfe der Methodenbetrachtung (Kap. 2.4.1) und dem Methodenvergleich (Kap. 2.4.2) gewonnenen Erkenntnissen lässt sich ableiten, dass für die Gebrauchstauglichkeitsevaluation in der Medizin eine Kombination der beiden Methoden Nutzertest mit realen Nutzern und Heuristische Evaluation geeignet ist, um möglichst viele Probleme zu identifizieren. Als problematisch wird bei der Heuristischen Evaluation teilweise angesehen, dass verwendete Heuristiken relativ allgemein gehalten und somit zu wenig spezifisch für Medizinprodukte sind [187]. Dutt et al. fanden heraus, dass Heuristiken teils nicht klar abgegrenzt und missverständlich sind [126]. In der Evaluationsphase dieser

Arbeit ist daher zu prüfen, ob die bekannten Heuristiken für den medizinischen Bereich ausreichend und anwendbar sind, oder ob spezielle Heuristiken für medizinische Geräte entwickelt werden sollten. Unbedingt soll in der Phase der Experimentenevaluation der vorhersehbare Missbrauch, den Forderungen der DIN EN IEC 60601-1-6 entsprechend, untersucht werden.

Die in der Heuristischen Evaluation evaluierten potentiellen Probleme einer Benutzerschnittstelle können mit einem Nutzertest verifiziert werden. Generell werden mit einem Nutzertest Leistungsmaße und subjektive Maße gemessen (vgl. Kap. 2.4.2). Leistungsmaße kommen bei der Effektivitäts- und Effizienzmessung zur Anwendung. Effektivität wird durch die Genauigkeit und Vollständigkeit der Aufgabenerfüllung definiert (vgl. Kap. 2.4). Um diese zu messen, müssen die Kriterien der erfolgreichen Zielerreichung operational definiert werden [113]. Hierbei kann sowohl die Qualität als auch die Quantität der Zielerreichung herangezogen werden. Die Übereinstimmung der Qualität des Ergebnisses mit dem festgelegten Kriterium kann als Grad der Genauigkeit bezeichnet werden. Als quantitative Zielerreichungsgröße kann der Prozentsatz der Zielerreichung als Vollständigkeitsmaß verwendet werden [113]. Diese Messung betrachtet nicht, wie die Aufgabenziele erreicht wurden, sondern nur den Umfang, mit dem sie erzielt worden sind. Als generell messbar gelten prozentuale Aufgabenerledigung, Fehlerhäufigkeit und Häufigkeit der Unterstützung. Effizienz hingegen betrachtet den im Verhältnis zur Effektivität eingesetzten Aufwand, mit dem ein Benutzer ein bestimmtes Ziel erreicht (vgl. Kap 2.4) [5, 113]. Mögliche Aufwandsgrößen können hierbei psychische oder physische Beanspruchung, Zeit, monetäre Kosten, Materialkosten oder Personalkosten sein. Um die Effizienz zu messen, muss der Grad der erreichten Effektivität zum benötigten Aufwand ins Verhältnis gesetzt werden [23, 113].

Die Feststellung dieser Leistungsmaße ist relativ einfach, da es sich hier um quantitative Daten handelt. In der Praxis am häufigsten gemessen werden die Aufgabenzeit, die Aufgabenerfüllung und die Anzahl der Fehler in einer Aufgabe [270]. Hierfür müssen lediglich entsprechende Vorrichtungen getroffen werden, um diese Daten genau zu ermitteln. Schwieriger ist die Analyse der Daten, da in der Regel keine allgemeinen Standards oder Vergleichswerte existieren, die zeigen, ob die erhobenen Daten brauchbar oder unbrauchbar sind. In der Literatur gibt es hierfür verschiedene Ansätze: Der Testverantwortliche legt die Kriterien fest. Messungen von Experten, Kriterien von bereits durchgeführten Tests, zwi-

schen den Verantwortlichen vereinbarte Werte, Vergleiche zu bestehenden Systemen oder Performannewerte mit einem Prototyp gelten als Maßstab [238, 411]. Bisher ist noch keine vollkommen zufriedenstellende Lösung gefunden worden, um genaue Spezifikationen festzulegen [351]. Die derzeit sinnvollste Lösung ist, in Zusammenarbeit mit Expertennutzern optimale Leistungsmaße festzusetzen und entsprechende Abweichungen zu bewerten. Sollten Vergleichsgeräte vorhanden sein, so können auch Messungen mit diesen Geräten als Leistungsmaße einfließen.

Das Gebrauchstauglichkeitskriterium der Zufriedenstellung wird in der Regel durch subjektive Maße gemessen. Es beschreibt das Ausmaß, in dem Benutzer von Beeinträchtigungen frei sind, und ihre Einstellungen zur Nutzung des Produktes. Die Zufriedenstellung ist durch verschiedene Meßmethoden evaluierbar. Sie kann gemessen werden durch z.B.:

- den erfahrenen Grad der Beeinträchtigung,
- subjektive Wertschätzung des Produktes,
- Zufriedenheit mit der Nutzung,
- Akzeptanz der Beanspruchung bei Ausführung verschiedener Aufgaben mit dem Gerät, wobei bestimmte Ziele der Gebrauchstauglichkeit als erfüllt angesehen werden,
- Anzahl positiver und negativer Äußerungen während der Nutzung [113].

Diese subjektiven Maße sind im Vergleich zu den objektiven schwerer zu bestimmen. Wie aber bereits in Kapitel 2.4.1 beschrieben, existieren verschiedene Methoden, um diese quantitativ und qualitativ zu erfassen. Beispielsweise ist der ASQ (vgl. Kap. 2.4.1.2) eigentlich eine subjektive qualitative Methode; durch Verwendung einer Likert-Skala werden jedoch auch quantitative Daten erhoben. Subjektive qualitative Daten werden durch die Methode des lauten Denkens ermittelt. Objektive Aussagen können aus Beobachtungen der Gestik, Mimik, Körperhaltung, Abwesenheitsrate oder aus der Beobachtung physiologischer Reaktionen der Nutzer gewonnen werden. Für die Erhebung der quantitativen Daten sollten standardisierte Fragebögen verwendet werden. Zusätzlich sollten Aussagen während des Tests und abschließende Interviews ausgewertet werden, um Rückschlüsse auf die Nutzerzufriedenheit zu ziehen.

Allgemein sind sowohl Daten zu einzelnen Handlungen und Arbeitsschritten als auch Daten über die allgemeine Einstellung gegenüber dem Produkt zu generie-

ren. Hierzu könnte während des Tests nach jeder Aufgabe ein kurzer Fragebogen zu der Zufriedenheit mit der einzelnen Aufgabe eingesetzt werden. Hierfür eignet sich der in Anhang A beschriebene ASQ.

Nach dem Test sollte noch einmal die das gesamte Gerät betreffende Zufriedenheit überprüft werden. Zur Ermittlung der allgemeinen Zufriedenheit mit dem Produkt und zur Überprüfung der Einhaltung der in der DIN EN 9241-11 beschriebenen Dialogprinzipien können bei Bedarf, je nach Nutzungskontext des Gerätes, entsprechend dem Kapitel 2.4.1.2 die Fragebögen PSSUQ, SUMI, SUS, QUIS, ISONorm, ErgoNorm oder IsoMetrics herangezogen werden (vgl. Anhang A). Der Vorteil dieser Fragebögen liegt darin, dass sie standardisiert sind und sich in der Softwareergonomie bewährt haben. Sie müssen jedoch für den medizinischen Nutzungskontext und das zu untersuchende Gerät angepasst werden. Derartige standardisierte Fragebögen für den medizinischen Bereich existieren noch nicht.

#### Einfachheit der Nutzung

	Stimme gar nicht zu				Stimme voll zu	
Es war einfach, das System zu nutzen.	---	--	-	+	++	+++
Es war einfach, die benötigte Information zu finden.	---	--	-	+	++	+++
Die vom System angebotene Information war klar.	---	--	-	+	++	+++
Es ist stets verständlich, welche Taste ich verwenden muss.	---	--	-	+	++	+++

Abb. 38: Beispielfragen zur Nutzerzufriedenheit, adaptiert nach dem PSSUQ [243]

Zur genaueren Identifikation der einzelnen Probleme sind Benutzerinterviews und spezielle Fragen nach der Nutzerzufriedenheit zu einzelnen Arbeitsaufgaben und Aktionen sinnvoll (Abb. 39).

Gesamteindruck des Instruments

Sehr schlecht	-3	-2	-1	1	2	3	Sehr gut
---------------	----	----	----	---	---	---	----------

Etwas greifen oder fassen

Sehr schlecht	-3	-2	-1	1	2	3	Sehr gut
---------------	----	----	----	---	---	---	----------

Das Maulteil rotieren

Sehr schlecht	-3	-2	-1	1	2	3	Sehr gut
---------------	----	----	----	---	---	---	----------

Abb. 39: Einstufung der Zufriedenheit zu einzelnen Nutzeraufgaben

Wenn verschiedene Geräte verglichen werden, bietet sich ein Ranking an, bei welchem die einzelnen Geräte gegenübergestellt werden (Abb. 40).

<p><b>Handling</b> Beurteilen Sie die Griffe im Hinblick auf ihre Handhabung.</p>	<p><b>Forcefeedback</b> Beurteilen Sie die Handgriffe im Hinblick auf Ihren Tastsinn.</p>
<p>Bitte ordnen Sie nun die Handgriffe entsprechend Ihrer Vorstellung: 1= entspricht am ehesten meiner Vorstellung, 4=entspricht am wenigsten meiner Vorstellung. Zwei Handgriffe dürfen nicht die gleiche Position einnehmen.</p>	

**Abb. 40: Vergleichendes Ranking zur Ermittlung der Nutzerzufriedenheit**

Weiterhin muss die für das medizinische Arbeitssystem wichtige „Sicherheit“ gemessen werden. Diese ist jedoch kein spezielles Maß für Gebrauchstauglichkeit, sondern wird durch die Maße Effektivität, Effizienz und Zufriedenheit unter Betrachtung des speziellen Nutzungskontextes erreicht. Bisher gibt es im Methodenrepertoire des Usability Engineering keinen Ansatz zur Sicherheitsbewertung. In dieser Arbeit wird hierfür eine spezielle Methode entwickelt, um die Gebrauchstauglichkeit unter dem Gesichtspunkt der Sicherheit zu bestimmen (► Kap. 6).

Thesen:	
T.11	Aus Nutzeranalyse, Anforderungen an das Gerät und einer Expertenevaluation können Testkriterien abgeleitet werden.
T.12	Nutzertest und Heuristische Evaluation decken die meisten Probleme auf.
T.13	Die Heuristische Evaluation sollte für medizinische Geräte angepasst werden.

T.14	Die Heuristische Evaluation erfasst potentielle Probleme, welche im Gebrauchstauglichkeitstest überprüft werden können. Hierzu sollten entsprechende Aufgaben gewählt werden.
T.15	Zur Messung der Gebrauchstauglichkeit in der Medizin werden Leistungsmaße und subjektive Maße benötigt.
T.16	Leistungsmaße können in Zusammenarbeit mit Expertennutzern aufgestellt werden.
T.17	Für die Erhebung subjektiver Zufriedenheitsdaten können Fragebögen eingesetzt werden. Mit einem aufgabenspezifischen Fragebogen (z.B. ASQ), speziellen Fragen zu den einzelnen Hauptaufgaben an dem Gerät und der Erfassung des Gesamteindrucks ist eine Bewertung der Gebrauchstauglichkeit möglich.
T.18	Um die Gebrauchstauglichkeit in der Medizin zu bewerten, reichen die Maße Effektivität, Effizienz und Zufriedenheit nicht aus, es müssen zusätzlich die Aspekte Sicherheit bzw. Risiko betrachtet werden.

### 4.1.3 Die Evaluatoren

Eine wichtige Rolle bei der Gebrauchstauglichkeitsevaluation kommt dem Evaluator und seinen Fähigkeiten zu. Für die Standardprozedur sind vier Evaluatorengruppen zu nennen. In diesem Kapitel werden die Testleiter, welche die Gebrauchstauglichkeitsevaluation durchführen, und die beiden Evaluatorengruppen für die Expertenevaluation (Heuristische Evaluation) bzw. die Bewertung der Gebrauchstauglichkeit behandelt. Der vierten Gruppe der Probanden für den Nutzertest wird danach ein eigenes Kapitel gewidmet.

#### Der Testverantwortliche

Für die Qualität einer Gebrauchstauglichkeitsevaluation kommt dem Testverantwortlichen eine wichtige Rolle zu. Unumstritten ist, dass der Gebrauchstauglichkeitstest durch einen erfahrenen Testverantwortlichen geleitet werden muss. Im Verlauf der Evaluation kommt er mit verschiedenen Personengruppen unterschiedlichster Fachkenntnisse und Interessen in Kontakt und kommuniziert mit ihnen. Er koordiniert und führt Interviews, Gruppensitzungen und Kontextbegehungen durch. Diese Aufgaben setzen ein hohes Maß an Feinfühligkeit, Diplomatie und sozialer Kompetenz voraus [36]. Weiterhin ist die fachliche Kompetenz des Testverantwortlichen von Bedeutung. Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich versiert sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird [24, 36].



Medizinische Geräte sind vielfältig. Die Standardprozedur soll für sie allgemeingültig sein. Es wird nicht vorausgesetzt, dass der Evaluator alle Bereiche der Medizin beherrscht, jedoch sollte er genügend technisches Know-How besitzen, um die Grundlagen der medizinischen Geräte zu verstehen, sowie eine Bereitschaft und Fähigkeit zur interdisziplinären Arbeit. Wichtige Voraussetzungen für den Evaluator sind außerdem ausgeprägte Kenntnisse über die Methoden der empirischen Forschung sowie das Verständnis der Mensch-Maschine-Interaktion und statistischer Analysen. Zusätzlich sollte er über den Zugang zu einer speziellen Testeinrichtung und zu einem Expertennetzwerk verfügen.

Von weiterer besonderer Bedeutung sind Kenntnisse über die Methoden der Gebrauchstauglichkeitsevaluation und über potentielle Fehlerquellen bei der Evaluation. Verschiedene Studien zeigen, dass unterschiedliche Evaluatoren trotz gleicher Evaluationsmethoden bei einem System unterschiedliche Gebrauchstauglichkeitsprobleme aufdecken. In unabhängigen Studien wurde untersucht, in welchem Umfang internationale Usability Labore die gleichen Probleme eines Systems entdecken [214, 268, 270, 272, 269]. Hierbei gab es erhebliche Abweichungen sowohl in den Ergebnissen als auch in der Methode.

Molich et al. ließen drei kommerzielle Usability Labore dasselbe System untersuchen. Diese bekamen jeweils eine zweiseitige Beschreibung der Nutzergruppe und der Ziele des Tests. Es stellte sich heraus, dass 129 der insgesamt 141 beschriebenen Probleme lediglich nur einmal angezeigt wurden [268]. In einer zweiten Studie mit sechs Usability Laboren und einer genaueren Beschreibung der Nutzer und der Testziele wurden von 186 bekannten Problemen 147 nur einmal benannt [272]. Kessner et al. kommen zu ähnlichen Ergebnissen. Sechs Teams führten unabhängig voneinander einen Gebrauchstauglichkeitstest mit derselben Software durch. Von insgesamt 36 identifizierten Gebrauchstauglichkeitsproblemen wurde keines von allen Teams gleichermaßen erkannt, zwei nur durch fünf, vier durch vier und weitere sieben durch zwei Teams [214]. Ähnliche Ergebnisse sind bei Molich & Dumas [269] sowie bei Molich & Ede et al. [270] zu finden. Weiterhin stellten Hertzum & Jacobsen bei einer Untersuchung von elf Studien mit verschiedenen Evaluationsmethoden fest, dass verschiedene Evaluatoren mit denselben Methoden verschiedene Arten von Gebrauchstauglichkeitsproblemen entdecken [182].

Es kommt also zu erheblichen Unterschieden bei der Gebrauchstauglichkeitsevaluation durch verschiedene Institutionen. Hertzum & Jacobsen beschreiben dies

als den Evaluatoreffekt (engl.: evaluator effect). Als Gründe für diesen Effekt sehen sie:

- eine ungenaue Zielanalyse bei der Wahl der Aufgabenszenarien für den Test,
- eine unzureichende Evaluationsprozedur und
- eine unspezifische Wahl der Kriterien zur Problemeinstufung [182].

Um diesen Effekt zu vermeiden und somit die Bewertungsqualität zu steigern, schlagen Hertzum & Jacobsen vor, dass die Evaluatoren in systematischer Weise die Auswahl ihrer Aufgabenszenarien verifizieren. Diese systematische Analyse soll sicherstellen, dass alle relevanten Systemparameter für die Erstellung eines Aufgabenszenarios berücksichtigt werden. Die wichtigsten Systemparameter für einen Test sind die mit einem Risiko verbundenen und solche, bei denen eine häufige Anwendungsfrequenz angenommen wird. Weiterhin sollten zumindest bei der Auswahl der Testkriterien zusätzliche Experten hinzugezogen und die Evaluationsprozedur und die Problemkriterien genau überlegt werden [182]. Außerdem sollte der Versuchsleiter berücksichtigen, dass er durch ungeschicktes Verhalten, wie z.B. unbewusste oder zu frühe Hilfestellung bzw. Suggestion in der Evaluationssituation das Ergebnis verfälschen kann. Nur im Ausnahmefall, wenn sich die Versuchsperson explizit an den Versuchsleiter wendet und um Hilfe nachfragt, sollte er sich aktiv einschalten, da ein zu frühes Eingreifen seinerseits die Versuchsperson daran hindern könnte, die Aufgabe selbstständig zu beenden [173].

### **Die Experten der Heuristischen Evaluation**

Wie bereits in Kapitel 2.4.2 gezeigt, eignen sich für die Heuristische Evaluation besonders Doppelperten mit Erfahrungen in der Medizin und in der Mensch-Maschine-Kommunikation [184]. Aber auch Experten eines dieser Wissensgebiete generieren zufriedenstellende Ergebnisse [73, 74, 282]. Bei der Wahl der Experten sollte auf ein ausgewogenes Verhältnis der Wissensgebiete geachtet werden. Hierbei sind neben Experten der Mensch-Maschine-Kommunikation auch solche aus den Anwendungsbereichen des medizinischen Produktes und der Hygiene sinnvoll. Auf jeden Fall sollten mehrere Evaluatoren hinzugezogen werden. Nielsen geht davon aus, dass bereits vier Experten ca. 75% der Probleme einer Benutzerschnittstelle aufdecken (vgl. Kap. 2.4.1.1). Andererseits ist umstritten, ob es sich auch bei allen um wirkliche Nutzerprobleme handelt, da die Experten schließlich keine Nutzer sind. In der Standardprozedur wird daher die Heuristische Evaluation als Vorbereitung für den eigentlichen Nutzertest angewandt.

### Die Experten der Problembewertung

Wie bei der Heuristischen Evaluation sollten für die Bewertung der Gebrauchstauglichkeitsprobleme Doppel-, Gebrauchstauglichkeits- und Fachexperten gewählt werden [212]. Hierbei ist auf die Unabhängigkeit der Experten zu achten, damit sie die Benutzerschnittstelle neutral bewerten können. Ferner ist einzukalkulieren, dass bei rein subjektiven Beurteilungen die Beurteiler unterschiedliche Meinungen bezüglich Gebrauchstauglichkeit und Sicherheit haben können. Daher ist unbedingt Nielsens Rat [284] zu folgen, nicht auf das Urteil eines einzelnen Spezialisten zu setzen, sondern verschiedene unabhängige Gutachter zu befragen und dann einen Mittelwert zu bilden. Um die Reliabilität der Beurteilung zu erhöhen, sollte die Expertenanzahl hoch sein. Nielsen hält drei bis vier Bewerter für ausreichend [284]. Der Ablauf einer Bewertungsmethode für die Messung der Gebrauchstauglichkeit medizinischer Geräte und die Anzahl der hierfür notwendigen Experten werden im Rahmen dieser Arbeit untersucht.

Thesen:	
T.19	Die Standardprozedur hilft, den Evaluatoreffekt zu vermeiden, da sie strukturiert den Nutzungskontext und die Ziele und Aufgaben analysiert, verschiedene Evaluatoren mit einbezieht und einen standardisierten Evaluationsleitfaden vorgibt.
T.20	Die Heuristische Evaluation sollte mit Doppelsexperten und Fachexperten durchgeführt werden.
T.21	Die Bewertung sollte mit mehreren Doppel- und Fachexperten durchgeführt werden.

#### 4.1.4 Bestimmung der Stichprobe für den Nutzertest

Die ausgewählte Stichprobe für eine Gebrauchstauglichkeitsbewertung medizinischer Geräte soll möglichst repräsentativ für die Nutzergruppe (vgl. Kap. 1.2.2.4) in Bezug auf demografische Faktoren, wie beispielsweise Aufgabenwissen, Bildungsniveau, Alter und ggf. Plattformkenntnis sein [90, 97, 105, 124, 284, 287, 371], d.h. sie soll hinsichtlich Qualifikation und Vorkenntnissen bezüglich des zu bewertenden Produktes dem Nutzerprofil (vgl. Kap. 4.1.1.1) entsprechen. Bei der Auswahl sind Personen zu vermeiden, welche stark von dem Benutzerprofil abweichen. Trotzdem sollten bei der Wahl der Testpersonen nicht nur die derzeitigen Repräsentanten der Nutzergruppen bzw. Nutzerexperten herangezogen werden, sondern auch zukünftige Nutzer und Novizen (Anfänger). Hierzu zählen beispielsweise Auszubildende, Studenten, Abteilungswechsler oder Nutzer aus

anderen Aufgabenbereichen [124, 287, 346]. Eine Studie von Molich et al., die neun Gebrauchstauglichkeitsstudien betrachtet, zeigt, dass in der Praxis meist ein ausgewogenes Verhältnis zwischen Laien und Experten besteht [270]. Diese Zusammensetzung hat den Vorteil, dass sie zum größten Teil die Eigenschaften der Nutzergruppe aufweist, aber dennoch unvoreingenommen an das Produkt herangeht. Betrachtet man eine Studie von Matern et al., so ist es gerade bei medizinischen Geräten im OP wichtig, dass auch bei einer Evaluation unerfahrene Nutzer als Stichprobe eingesetzt werden, denn 58,8% von 410 befragten Chirurgen und 40,3% von 186 befragten OP-Schwestern fühlten sich nicht ausreichend für die Bedienung der medizinischen Geräte geschult [256]. Somit ist davon auszugehen, dass es im medizinischen Kontext immer wieder zu Situationen kommen kann, in welchen unerfahrene Nutzer medizinische Geräte bedienen müssen.

Sind bei der Evaluation zusätzlich körperliche Eigenschaften zu berücksichtigen, so legt die DIN 33419 ein Verfahren zur Zusammenstellung einer Personengruppe fest, welche in ihren körperlichen Eigenschaften repräsentativ für die erwartete Gesamtheit aller zukünftigen Benutzer des zu prüfenden Gegenstandes ist. Sie definiert die Prüfpersonen und deren Auswahlgesichtspunkte nach speziellen Basiskriterien (Geschlecht, Größe, Alter, Gewicht etc.) [90]. Für eine möglichst variable Erfassung sind die Testnutzer entsprechend den relevanten Kriterien und der Prüfaufgabe auszuwählen. Für ein Prüfkriterium sind jeweils mindestens drei Prüfpersonen notwendig. Diese dürfen jedoch auch mehrere Kriterien abdecken.

### **Stichprobengröße**

Bei einem Nutzertest tritt die elementare Frage auf, wie viele Testnutzer benötigt werden. In der Literatur findet man verschiedene Ansichten [124, 138, 152, 181, 240, 288, 394]. Nielsen & Landauer berufen sich auf ein mathematisches Modell. Sie fanden heraus, dass die Zahl der Gebrauchstauglichkeitsschwachstellen, die in einem Test mit „n“-Benutzern entdeckt werden, sich annähernd durch die Formel

$$\text{Gebrauchstauglichkeitsschwachstellen} = N(1-(1-\lambda)^n)$$

berechnen lässt [288]. Hierbei drückt „N“ die Gesamtzahl der Schwachstellen des Gerätes und  $\lambda$  den Anteil der entdeckten Schwachstellen (Problemfindungsrate<sup>36</sup>) durch den einzelnen Nutzer aus. Anhand verschiedener Studien fanden sie heraus, dass ein typischer Wert von  $\lambda$  bei 31% liegt. Sie sehen bei der Gebrauchstauglichkeitsuntersuchung eine angemessene Kosten-Nutzen-Relation bei einer Teilnehmerzahl von drei bis vier gegeben. Faulkner geht hingegen von einer größeren Evaluatorengruppe in einem derartigen Test aus. In einer 2003 durchgeführten Studie prüfte sie in einem Gebrauchstauglichkeitstest mit 60 Nutzern Niensens Theorie. Hierbei wählte sie wahllos jeweils fünf Testpersonen aus und betrachtete den prozentualen Anteil der insgesamt gefundenen Fehler. Dieser lag zwischen 55 und 99% bei Gruppen zu je fünf Nutzern (im Durchschnitt 85,55%). Bei zehn Nutzern wurden mindestens 82% (im Durchschnitt 94,69%), bei 15 mindestens 90% (im Durchschnitt 97,05%) und bei 20 wurden mindestens 95% (im Durchschnitt 98,4%) der Probleme entdeckt. Einen eindeutigen Vorschlag gibt Faulkner zwar nicht, aber man kann bei diesem Ansatz von einer angemessenen Aufwand-Nutzen-Relation durch den Einsatz von ca. 15 Testnutzern ausgehen [138].

Bei Dumas hängt die Größe einer Stichprobe von mehreren Faktoren ab: a) wieviele Testnutzeruntergruppen<sup>37</sup> man benötigt, um die Ziele zu erreichen, b) wieviel Zeit und Geld man für den Test zur Verfügung hat und c) wie wichtig die statistische Signifikanz der Ergebnisse ist [124]. Er schlägt für einen typischen Test 6-12 Teilnehmer in zwei bis drei Untergruppen vor. Lewis gibt die Zahl der Testnutzer in Abhängigkeit von dem Produkt als weit höher an [240].

Was sagen also die variierenden Meinungen bezüglich der Testnutzeranzahl für einen Gebrauchstauglichkeitstest aus? Wenn 15 Testnutzer benötigt werden, um 90% der Probleme aufzudecken, warum werden dann in der Regel sechs Nutzer empfohlen [11, 284]? Die optimale Anzahl von Probanden hängt in hohem Masse davon ab, welches Produkt getestet wird und welche Ziele mit dem Test jeweils verfolgt werden. Prinzipiell ist es sinnvoll, in der Produktentwicklung den für die Gebrauchstauglichkeitsevaluation bereitgestellten Etat für einen iterativen Prozess aus mehreren kleineren formativen Tests mit einer geringeren Stichprobe zu verwenden. Bei einem Gebrauchstauglichkeitstest, der als Bewertung dienen soll,

---

<sup>36</sup> Mit Hilfe einer Formel nach Lewis kann die Problemfindungsrate ( $p$ ) unter Einbeziehung eines Good-Turing-Schätzers berechnet werden [245].

<sup>37</sup> Nutzergruppen welche bestimmte Charakteristiken teilen, z.B. erfahrene Nutzer vs. unerfahrene [124]

mit dem Ziel, möglichst viele Probleme aufzudecken, um die Validität zu sichern, sollte dagegen eine gründliche summative Evaluation (vgl. Kap. 2.4.1) angewandt werden, um möglichst viele Schwachstellen zu entdecken. Trotzdem sollte nicht zwingend davon ausgegangen werden, dass 15 Testnutzer benötigt werden, um diese Bewertung durchführen zu können. In die Entscheidung sollte zusätzlich der Gedanke einfließen, dass das Produkt von mehreren Nutzergruppen verwendet werden könnte und somit auch an verschiedenen Nutzergruppen getestet werden müsste. Für medizinische Produkte sollte zusätzlich der Aspekt von Sicherheit und Risiko in die Festlegung der Testnutzeranzahl einfließen. In sicherheitskritischen Bereichen, wie z.B. in der Medizin, sollten möglichst alle potentiell sicherheitskritischen Fehlerquellen aufgedeckt werden, selbst bei hohen Kosten [350]. Ein möglicher Vorschlag für die Stichprobengröße unter Berücksichtigung der einzelnen Faktoren findet sich in Tab. 14. Diese veranschaulicht die ökonomische Rechtfertigung mit den für die Gebrauchstauglichkeit wichtigen Faktoren, welche die Gruppengröße bei der Evaluation medizinischer Geräte beeinflussen können.

	5 Nutzer		30 Nutzer
Komplexität des Produktes	gering	→	hoch
Unterschiedliche Nutzergruppen	eine	→	viele
Risiko	marginal	→	hoch

Tab. 14: Anzahl der Testnutzer in Abhängigkeit von der Produktkomplexität, der Homogenität der Nutzergruppe und von dem mit dem Gerät verbundenen Risiko.

Thesen:	
T.22	Die Stichprobe sollte ein ausgewogenes Verhältnis von Laien zu Fachnutzern aufweisen.
T.23	Für die Stichprobe sind Probanden zu nehmen, welche dem Nutzerprofil entsprechen.
T.24	Die Anzahl der Nutzer muss bei der Gebrauchstauglichkeitsprüfung in der Medizin höher als 6 sein. Die Problemfindungsrate sollte bei mindestens 90% liegen.
T.25	Mit einer Stichprobengröße von 9 Nutzern werden mehr als 80% der Probleme gefunden. Hiermit ist meist eine Bewertung möglich.

#### 4.1.5 Szenarien und Testaufgaben

Für einen Nutzertest ist es primär wichtig, Arbeitsaufgaben zu erstellen, welche die Testnutzer abarbeiten. In diesem Zusammenhang sind Szenarien von besonderer Bedeutung. Der Begriff Szenario ist vielseitig, da er in verschiedenen Dis-

ziplinen unterschiedlich eingesetzt wird [41, 66]. Dies gilt auch für den Bereich der Mensch-Maschine-Kommunikation. Bei Carrol & Rosson findet man beispielsweise sieben verschiedene Bedeutungen [55].

Häufig treten Szenarien im Rahmen von Bewertungsmethoden der Interaktionsarchitektur auf. In diesem Zusammenhang wird unter einem Szenario die kurze Beschreibung einer einzigen Interaktion eines Betroffenen oder einer Interessengruppe (z.B. Kunden, Pflegepersonal etc.) mit einer Anwendung verstanden [41]. Szenarien werden hier bei der Beschreibung von Anwendungsfällen verwendet. Dementsprechend kann ein Szenario mit dem Wort „Ablauf“ beschrieben werden. Es stellt eine Instanz des Anwendungsfalls dar. Oestereich definiert ein Szenario als eine „spezifische Folge von Aktionen“, wie beispielsweise einen konkreten Ablaufpfad in einem Anwendungsfall [298]. Glinz bezeichnet ein Szenario als eine geordnete Menge von Interaktionen zwischen Partnern [153]. In der Regel ist dies die Interaktion zwischen einem System und einer Menge systemexterner Akteure. Man kann hierbei zwischen Kontext- und Anwendungsszenarien unterscheiden. Die für die Szenarienerstellung notwendigen Informationen können aus der Anforderungsanalyse (vgl. Kap. 4.1.1) oder aus gesonderten Erhebungen gewonnen werden.

**Kontextszenarien** kommen aus dem Bereich der Softwaregestaltung und sind „episodische Beschreibungen von Aufgaben und Tätigkeiten in ihrem Kontext“ [77]. Sie können aber auch auf andere Produkte übertragen werden. Das Ziel eines Kontextszenarios ist es, Anforderungen aus dem Nutzungskontext (vgl. Kap. 4.1.1) mit Hilfe der Dialogprinzipien [114] abzuleiten. Im Fokus dieser, in natürlicher Sprache verfassten Beschreibung steht die Darstellung der Zielerreichung eines Nutzers, ohne dass auf eine spezielle Produktnutzung eingegangen wird. Die Szenarien sollten daher nicht durch spezielle Produktfeatures eingeschränkt werden, sondern primär auf die Ziele der Nutzer fokussieren. Kontextszenarien werden in der Vorbereitung eines Entwicklungsprojektes oder anlässlich einer Softwareprüfung eingesetzt [77]. Die Datenerhebung erfolgt in der Regel durch Nutzerinterviews. Im Prüfhandbuch der DATech wird ein Leitfaden für die Erstellung von Kontextszenarien beschrieben [77].

**Anwendungsszenarien**, auch Nutzungsszenarien oder Gebrauchsszenarien genannt, meist besser bekannt unter dem englischen Ausdruck „Use-Scenario“,

erfassen die Interaktion des Benutzers. Sie sind episodische Darstellungen von Handlungen (wirklichen Aufgaben), welche Anwender und Betreiber mit dem Gerät durchführen [77, 97]. Bei der Anforderungsanalyse (vgl. Kap. 4.1.1) kann durch Anwendungsszenarien die Interaktion der Primäraufgabe beschrieben werden, um diese dann mit realisierten Merkmalen der Benutzerschnittstelle zu vergleichen. In der Produktevaluation wird die Interaktion in ihren Grundzügen erfasst, um diese dann in einem Test zu überprüfen. Die Szenarien können z.B. in Form von Anwendungsfällen beschrieben werden.

Diese Arten von Szenarien werden häufig in Gestalt kleiner Handlungsgeschichten als Methode eines benutzerzentrierten Entwicklungsprozesses eingesetzt. Diese meist personalisierten, fiktionalen Geschichten über eine Produktnutzung in einem bestimmten Nutzungskontext fördern die Kommunikation zwischen Benutzer und Entwicklern. Die Ergebnisse können direkt in die Produktentwicklung einfließen und als Evaluationsgrundlage dienen [56, 160, 284, 314]. Für Nielsen sind diese Szenarien bereits erste Prototypen [284]. Um Anwendungsszenarien zu erstellen, müssen Ziele, Aufgaben und Handlungen zur Zielerreichung, die Arbeitsumgebung sowie die Anwender selbst betrachtet werden. Hierzu können Ergebnisse aus der Anforderungsanalyse verwendet und bei Bedarf neue Daten erhoben werden. Die DIN EN IEC 60601-1-6 und DIN EN IEC 62366 legen fest, dass die Gebrauchsszenarien durch den Hersteller in der Spezifikation der Gebrauchstauglichkeit beschrieben werden müssen [97], um die Gebrauchstauglichkeit des Gerätes zu bestimmen und zu prüfen.

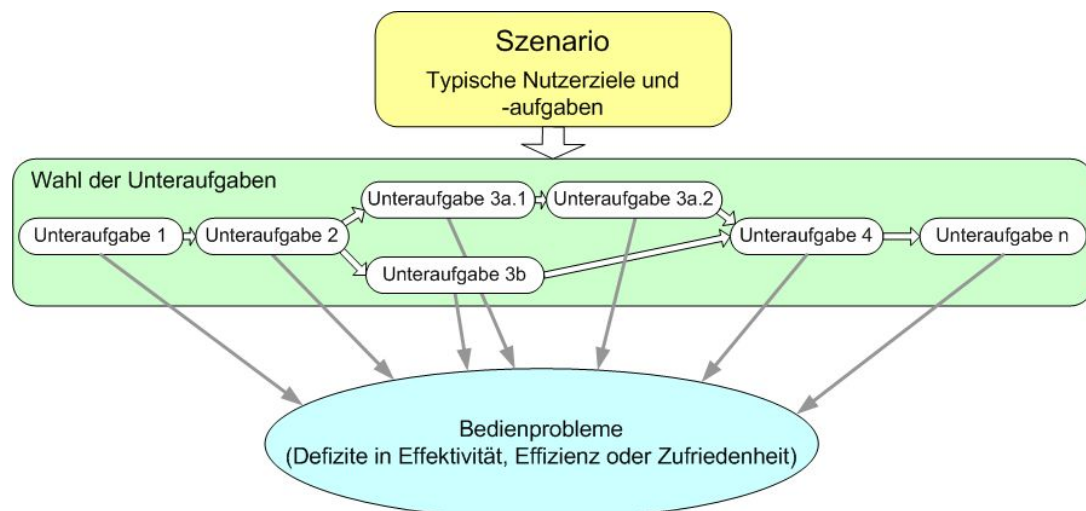
Aus den Anwendungsszenarien, den Kontextszenarien und den Informationen der Anforderungsanalyse werden die für die Evaluation notwendigen repräsentativen Testaufgaben abgeleitet. Generell ist es sinnvoll, alle möglichen Aufgaben mit dem Gerät zu evaluieren. Bei sehr komplexen Produkten, bei denen zu viele Arbeitsaufgaben erkannt werden, um sie in einer angemessenen Zeit zu testen, muss entschieden werden, welche Arbeitsaufgaben die wichtigsten für den Gebrauchstauglichkeitstest sind. Bei der Selektion sind spezielle Entscheidungskriterien zu berücksichtigen [125].

Um eine möglichst hohe Reliabilität zu gewährleisten und den in Kapitel 4.1.3 beschriebenen Evaluatoreffekt zu vermeiden, sollten in jedem Test möglichst ähnliche Aufgaben evaluiert werden. Testaufgaben sollten hauptsächlich Primäraufgaben, Aufgaben in Notsituationen sowie Sekundäraufgaben, welche in der



Anforderungsanalyse oder der Heuristischen Evaluation als potentiell problematisch oder risikoträchtig identifiziert wurden, enthalten. Zusätzlich sollte auch, entsprechend den Forderungen der DIN EN IEC 60601-1-6, ein vernünftigerweise vorhersehbarer Missbrauch betrachtet werden [97]. Bei der Wahl der Testaufgaben ist darauf zu achten, dass sie in einem zeitlichen Limit zu erfüllen sind und den Nutzer nicht durch ihr Ausmaß in überhöhtem Maße beanspruchen. Dennoch dürfen sie nicht trivial wirken.

Nach Festlegung der in der Gebrauchstauglichkeitsevaluation geprüften Arbeitsaufgaben und Nutzerziele werden diese in Aufgabenszenarien übertragen. Hierbei beschreibt jedes Szenario eine Aufgabe, welche von den Versuchspersonen zu bearbeiten ist. Bei der Wahl der Reihenfolge dieser Aufgaben soll auf Logik und Sinnhaftigkeit geachtet werden [173]. Diese Szenarien stellen somit typische Arbeitsaufgaben dar, mit denen die Nutzer ein Ziel erreichen wollen. Die Aufgaben können in mehrere Unteraufgaben unterteilt werden, wobei es möglich ist, dass unterschiedliche Nutzer unterschiedliche Unteraufgaben zur Problemlösung wählen. Abb. 41 veranschaulicht diesen Zusammenhang:



**Abb. 41: Szenario und Aufgaben einer Gebrauchstauglichkeitsprüfung**

Die dem Nutzer gestellten Arbeitsaufgaben sollten auf Nutzerziele ausgerichtet und dem realen Nutzungskontext des Gerätes entnommen sein. Diese Aufgaben werden normalerweise dem Testnutzer in Schriftform vorgelegt, können aber auch in einem Dialog vermittelt werden. Um sie realistischer zu gestalten, werden sie in Szenarien beschrieben. In einem solchen Szenario gibt man dem Nutzer in seinen Worten mit Hilfe einer kleinen Geschichte ein Ziel und weitere Informationen vor, die er für die Zielerreichung benötigt. Genauere Informationen über die

Aufgabenerledigungen werden jedoch nicht gegeben [124]. Ein gutes Szenario ist kurz, für den Nutzer verständlich, gibt genügend Informationen, ist weder humorvoll noch frivol und direkt mit den Aufgaben und dem zu evaluierenden Evaluationsziel des Testszenarios verbunden [124, 270, 284]. Ein entsprechendes Beispiel ist in Abb. 42 veranschaulicht.

<p><b>Evaluationsziel:</b> Können Nutzer bei einem Hochfrequenzchirurgie-Gerät einen Fußschalter zuweisen? Werden sie bei der Zuweisung durch das Gerät ausreichend unterstützt? Werden evtl. auftretende Hinweismeldungen verstanden?</p> <p><b>Testaufbau:</b> Bei dieser Aufgabe ist ein Fußschalter angeschlossen, jedoch noch nicht dem Instrument zugewiesen. Der Gerätemonitor zeigt das Hauptmenü.</p> <p><b>Testbeschreibung:</b> Es soll herausgefunden werden, wie ein Fußschalter dem Instrument zugewiesen wird.</p> <p><b>Testnutzer:</b> OP-Schwestern</p> <p><b>Szenario:</b> Der leitende Chirurg möchte eine Gefäßversiegelung mit einer Bipolarpinzette durchführen. Er möchte für die Aktivierung der Pinzette einen Fußschalter verwenden. Er bittet Sie, die entsprechenden Einstellungen vorzunehmen.</p>
--

**Abb. 42: Beispiel eines Aufgabenszenarios**

<b>Thesen:</b>	
T.26	Anwendungsszenarien helfen bei der Definition des Nutzungskontextes und der Erstellung der Testaufgaben.
T.27	Aufgabenszenarien vermitteln Nutzern in natürlicher Weise das Aufgabenziel und Informationen für die Aufgabe.
T.28	Aus den Anwendungsszenarien, den Kontextszenarien und den Informationen der Anforderungsanalyse können die für die Evaluation notwendigen repräsentativen Testaufgaben abgeleitet werden.
T.29	Testaufgaben sollen hauptsächlich Primäraufgaben enthalten. Es sollen aber auch Sekundäraufgaben und Aufgaben in Notsituationen gestellt werden sowie Aufgaben, welche in der Anforderungsanalyse oder in einem Expertenreview als potentiell problematisch oder risikoträchtig identifiziert wurden. Zusätzlich soll auch, entsprechend den Forderungen der DIN EN IEC 60601-1-6, ein vernünftigerweise vorhersehbarer Missbrauch betrachtet werden.

#### 4.1.6 Testaufbau und Testumgebung

Evaluationsverfahren können in der realen Nutzungsumgebung oder in einer Laborumgebung durchgeführt werden. Nutzertests sollten möglichst die reale Nutzungssituation wiedergeben. Die Laborumgebung hat gegenüber der realen Nutzungsumgebung den Vorteil, dass die Untersuchungsbedingung zu einem hohen Grad kontrollierbar und somit reproduzierbar ist, was zur Erfüllung der

Testgütekriterien Validität und Reliabilität beiträgt. Auch die Nutzung technischer Hilfsmittel wie Aufnahme- und Analysesysteme ist hier möglich. Das Labor muss so ausgestattet sein, dass der Testnutzer die Testumgebung als real wahrnimmt. Auch sollte die notwendige Beobachtungssituation die Anwender in dieser Umgebung nicht stören. Trotzdem sollten die Vorzüge einer Laborumgebung nicht verloren gehen [117, 315]. In einer Laborumgebung können Anwenderszenarien erzeugt werden, die sicherheitsrelevant sind, jedoch in der Realität nur selten auftreten und aus ethischen Erwägungen nicht systematisch hervorrufbar und deshalb auch nicht in der realen Arbeitsumgebung zu evaluieren sind. Gerade im hochsensiblen medizinischen Bereich kann eine beobachtende Person bei solchen Extremsituationen in der realen Nutzungsumgebung störend und sogar risikofördernd wirken. Daher sind Gebrauchstauglichkeitstests in einer realitätsgetreuen Laborumgebung, in der die Anwendungsbedingungen kontrollierbar sind und keine Patienten gefährdet werden, durchzuführen [149]. Diese Usability-Labore sind für die Datenerhebungsverfahren technisch optimal ausgerüstet. Störvariablen können eliminiert oder zumindest kontrolliert werden, wodurch ein Vergleich zwischen verschiedenen Versuchspersonen und Produkten möglich ist. Für spezielle Nutzungsumgebungen gibt es verschiedene Labortypen. Grundsätzlich gleiche Komponenten sind: 1. Ein Testarbeitsplatz, der entsprechend dem zu untersuchenden Objekt ausgerüstet ist, 2. ein von diesem Raum abgetrennter Kontrollraum, von dem aus die Versuchsperson beobachtet werden kann, und 3. die Aufzeichnungstechnik, welche in Bild und Ton die zu testenden Abläufe und die Reaktionen der Versuchspersonen dokumentiert.

Ein spezielles Usability-Labor für medizinische Produkte ist der Tübinger „Experimental-OP“<sup>38</sup> (Abb. 43). In diesem Labor wurde ein medizinischer Arbeitsbereich aufgebaut, mit dem Ziel, diesen zu optimieren, Kosten einzusparen und ihn sicher zu gestalten. Aktuelle und zukünftige medizinische Entwicklungen können hier integriert und hinsichtlich ihrer Gebrauchstauglichkeit überprüft werden.

---

<sup>38</sup> Weitere Informationen unter [www.wwH-c.com](http://www.wwH-c.com)



**Abb. 43: Der Experimental-OP Tübingen beinhaltet neben zwei OPs mit den notwendigen Nebenräumen Ein- und Ausleitung, Lager für Geräte und Sterilgut auch Aufwachräume, Intensivzimmer, Schockraum und Eingriffsraum.**

<b>These:</b>	
T.30	Gebrauchstauglichkeitsuntersuchungen in der Medizin müssen in einer Laborumgebung durchgeführt werden, in welcher der Nutzungskontext simuliert werden kann.

#### 4.1.7 Report

Viele Abschlussberichte von Gebrauchstauglichkeitsevaluationen sind selbst nicht gebrauchstauglich. Häufige Fehler werden von Dumas & Redish, sowie von Molich et al. dargestellt [124, 270]. In einigen Fällen ist beispielsweise der Bericht zu lang, in anderen werden zu viele Fehler dargestellt. Weitere Mängelpunkte sind: Es existieren keine Kurzfassung, keine Klassifikation der Probleme nach Schweregrad, keine Aussage über die Zahl der Nutzer, die mit dem Problem konfrontiert wurden, keine Positivaussagen oder es existiert nur ein unattraktives resp. unprofessionelles Layout bzw. eine unklare oder vage Problembeschreibung.

Um Fehler zu vermeiden und dem Testgütekriterium der Reliabilität zu genügen, ist entsprechend Molich et al. ein standardisierter Bericht notwendig [270]. Seit 2006 wird durch den internationalen Standard ISO 25062 ein allgemeiner Standard für Gebrauchstauglichkeitsberichte festgelegt [200]. Dieser baut auf der Definition von Gebrauchstauglichkeit nach der DIN EN 9241-11 auf [113] und standardisiert die Informationen, welche bei einem Gebrauchstauglichkeitstest für Softwareprodukte erfasst und dokumentiert werden sollen. Der Detaillierungsgrad dieses Standards erlaubt der durchführenden Organisation und auch jeder ande-

ren eine Wiederholung des Prüfverfahrens, womit dieser Standard zur Reliabilität der Gebrauchstauglichkeitstestreihe beiträgt. Die wichtigsten Elemente des Berichtes sind: demografische Daten der Nutzer, Aufgabenbeschreibungen, Testkontext mit verwendetem Equipment, Testumgebung und Testunterlagen sowie die verwendeten Maße zur Evaluation. Die Hauptabschnitte des Berichtes und eine genaue Checkliste sind der ISO 25062 zu entnehmen [200].

Die ISO 25062 ist zwar für Softwareprodukte vorgesehen, eignet sich in ihren Grundzügen aber auch für medizinische Geräte. Daher kann der in dieser Norm beschriebene Standard mit geringen Änderungen auch im medizinischen Bereich Anwendung finden. Um sie für den medizinischen Kontext anzupassen, muss der sicherheitskritische Aspekt beachtet werden. Abb. 44 gibt ein Beispiel, wie diese Berichtvorgabe für den medizinischen Kontext angepasst werden kann. Neben der Integration einer Sicherheitsbetrachtung wurde zusätzlich die logische Reihenfolge verändert und weitere Punkte hinzugefügt, um die Verständlichkeit des Berichtes zu optimieren.

- **Deckblatt**
- **Zusammenfassung**
- **Einleitung**
  - Produktbeschreibung
  - Testziele
- **Testkriterien**
- **Gebrauchstauglichkeitsmaße (Effektivität , Effizienz u. Zufriedenheit)**
- **Methode**
  - Kontext der Produktverwendung in dem Test (Benutzerprofil, Aufgaben, Umgebungsanalyse, Geräteumgebung des Teilnehmers, Sicherheitsbetrachtung/Sicherheitsanalyse)
  - Versuchsplan (Ablauf, Testmethoden, Testaufgaben und Aufgabenszenarien, Prüfziele des Gebrauchstauglichkeitstests, Testumgebung, Testdesign u. -prozedur, Allgemeine Testanweisung, Aufgabenanleitung, Testwerkzeuge, z.B. Hardware, Fragebögen, Teilnehmer)
- **Evaluation**
  - Heuristische Evaluation
  - Gebrauchstauglichkeitstest
- **Ergebnisse**
  - Datenanalyse
  - Problem- und Fehleranalyse
  - Sicherheitsbetrachtung und Problembewertung
  - Präsentation der Ergebnisse (Leistungsergebnisse , Zufriedenheitsergebnisse , Ergebnisse der Sicherheitsbetrachtung)
  - Anhänge

**Abb. 44: Abschnitte eines standardisierten Berichtes für eine Gebrauchstauglichkeitsevaluation medizinischer Geräte**

<b>These:</b>	
T31	Der für Softwareprodukte standardisierte Gebrauchtauglichkeitsbericht kann in einer angepassten Form für Medizinprodukte verwendet werden.

#### 4.1.8 Qualität des standardisierten Verfahrens

Eine Evaluation muss die klassischen Testgütekriterien Validität, Reliabilität und Objektivität erfüllen [91, 36, 168, 199, 265]. Dies kann durch Wiederholbarkeit, Reproduzierbarkeit, Unparteilichkeit und Sachlichkeit erreicht werden. Gleiches gilt auch für die Evaluation der Gebrauchstauglichkeit. Laut Richter & Flückig können diese Testgütekriterien bei Gebrauchstauglichkeitstests weitestgehend eingehalten werden [336]. Im Folgenden werden die einzelnen Testgütekriterien besonders in Bezug auf die Standardprozedur erläutert.

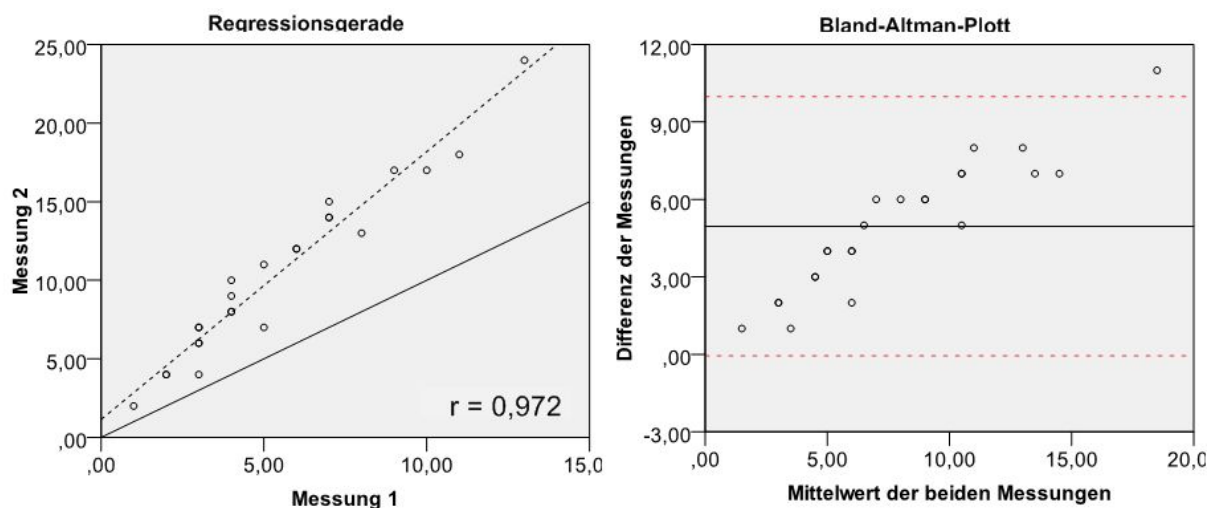
**Objektivität** drückt die Unabhängigkeit der mit einem Verfahren erzielten Ergebnisse vom Auftragnehmer und seinen Mitwirkenden, von den situativen Umständen sowie von weiteren irrelevanten Einflüssen aus [91]. Ein Test ist objektiv, wenn verschiedene Testanwender bei denselben Personen zu den gleichen Resultaten gelangen. Somit soll ein Test unabhängig vom Testanwender sein. Man unterscheidet hierbei:

- Durchführungsobjektivität: Das Ergebnis soll vom Untersuchungsleiter unbeeinflusst sein.
- Auswertungsobjektivität: Die Vergabe von Testpunkten muss vom Evaluator unbeeinflusst sein.
- Interpretationsobjektivität: Individuelle Deutungen dürfen in die Interpretation eines Testwertes nicht einfließen.

Objektivität wird dadurch erreicht, dass standardisierte Elemente verwendet werden. Dies gilt besonders für Fragebögen, aber auch die Aufgabenwahl kann zum größten Teil dadurch standardisiert werden, dass eine klar festgelegte Vorgehensweise definiert wird. Durch die Standardprozedur wird ein Standard festgelegt, der eine höhere Objektivität in diesem Bereich sicherstellen soll. Die Durchführungs- und Auswertungsobjektivität werden durch das standardisierte Vorgehen und die Verwendung standardisierter Methoden gefördert, die Interpretationsobjektivität dadurch, dass ermittelte Ergebnisse auf eine festgelegte Klassifizierung übertragen werden, welche sich an den in den einschlägigen Normen festgelegten Gebrauchstauglichkeitsmaßen orientiert [97, 104, 113].

Die Bestimmung von Objektivität erfolgt im Allgemeinen über den Korrelationskoeffizienten ( $r$ )<sup>39</sup>. Von Objektivität kann gesprochen werden, wenn die Korrelation der Ergebnisse zwischen verschiedenen Testleitern hoch ist. Bös gibt für den Korrelationskoeffizienten folgende Richtwerte an:  $r > 0.9$  = ausgezeichnet;  $0.8-0.9$  = sehr gut;  $0.7-0.8$  = annehmbar;  $0.6-0.7$  = mäßig;  $< 0.6$  = gering [37].

Bland & Altman geben jedoch zu bedenken, dass eine hohe Korrelation nicht bedeutet, dass zwei Messmethoden gleich sind, sondern lediglich, dass die Werte voneinander abhängig sind (Abb. 45). Wenn beispielsweise das zweite Messverfahren ein Vielfaches des ersten misst, so ist die Korrelation immer noch Eins. Daher schlagen sie ein grafisches Verfahren vor, um die Gleichheit zwischen zwei Messverfahren zu beschreiben. Hierzu werden die gemittelten Ergebnisse zweier Messmethoden (x-Achse) gegen ihre Messwert-Differenzen (y-Achse) aufgetragen. Der Mittelwert der Messwert-Differenzen ( $\bar{d}$ ) wird auch als systematischer Fehler oder Bias bezeichnet. Bei Normalverteilung liegen 95% der Werte innerhalb der 1,96-fachen Standardabweichung ( $s$ ), also zwischen  $\bar{d} - 1,96s$  und  $\bar{d} + 1,96s$ . Je kleiner dieser Bereich ist, desto mehr nähern sich die Messmethoden einander an [29]. Sofern dieser Unterschied zwischen  $\bar{d} \pm 1,96s$  innerhalb eines annehmbaren Rahmens ist, sind die beiden Messverfahren vergleichbar.



**Abb. 45: Regressionsgerade und Bland-Altman-Plott. Obwohl eine Korrelation von  $r=0,97$  der beiden Messungen vorliegt, zeigen sowohl die Regressionsgerade als auch der Bland-Altman-Plott, dass die Messungen nicht vergleichbar sind.**

<sup>39</sup> Dieser Koeffizient wird zur Überprüfung verwendet, ob zwei Methoden das Gleiche messen. Er ist ein Maß für den Zusammenhang zwischen zwei Variablen und kann Werte zwischen -1 und +1 annehmen.

**Reliabilität** ist der Grad der Genauigkeit, mit der das Verfahren das gemessene Merkmal erfasst [91]. Folglich müssen unter den gleichen Bedingungen die gleichen Ergebnisse reproduziert werden können. Grundsätzlich ist eine perfekte Reliabilität in der Praxis nicht zu erreichen [36], da sich gewisse Fehler nie ausschließen lassen. Ziel ist es, den Messfehler möglichst gering zu halten.

Oft wird bei Gebrauchstauglichkeitsuntersuchungen die Reliabilität als problematisch angesehen [214, 165, 270]. Verschiedene Studien zeigen, dass in der Praxis große Differenzen in den erkannten Problemen, Methoden, Aufgaben und dem Abschlussreport feststellbar sind [214, 270]. Hierbei ist der bereits in Kapitel 4.1.3 erwähnte Evaluatoreffekt zu nennen. Eine große Schwierigkeit liegt auch darin, dass meist sehr unterschiedliche Nutzer und Nutzergruppen vorliegen. Bei einem Nutzertest ist es nicht ungewöhnlich, dass der schnellste Testnutzer zehnmals schneller seine Arbeitsaufgabe erfüllt als der langsamste und dass ein Viertel der schnellsten immer noch zweimal so schnell ist wie ein Viertel der langsamsten [284]. Nutzer mit unterschiedlichem Wissen und Erfahrungen finden häufig auch unterschiedliche Probleme. Bei jedem Gerät wird es Anwendungsschwierigkeiten geben, die fast mit jeder Testperson feststellbar sind, aber auch Probleme die nicht so einfach zu entdecken sind. Eine ausgewogene Wahl der typischen Testnutzergruppen ist daher von besonderer Wichtigkeit, und mit zunehmender Testteilnehmerzahl steigt die Reliabilität [284].

Weitere Schwierigkeiten für eine unzureichende Reliabilität liegen darin, dass die Testmethoden und Testumgebungen variieren, unterschiedliche Testaufgaben gewählt werden, keine ausreichende Aufgaben- und Nutzungskontextanalyse durchgeführt wird, keine standardisierten Klassifizierungs- und Bewertungsmöglichkeiten für die Gebrauchstauglichkeitsprobleme gewählt, Gebrauchstauglichkeitsprobleme nicht von anderen Problemen differenziert und unterschiedliche Berichte vorgelegt werden.

Zur Vermeidung dieser Problematik und zur Erreichung von mehr Reliabilität ist eine Standardisierung aller Elemente der Gebrauchstauglichkeitsevaluation unabdingbar. Hierzu gehören:

- Ermittlung der Nutzerziele, Arbeitsaufgaben und Aktionen (vgl. Kap. 4.1.1) sowie der Gebrauchstauglichkeitsziele in enger Zusammenarbeit mit Anwendern und Herstellern (vgl. Kap. 4.1.2),
- Kombination von geeigneten Testmethoden, um mögliche Probleme zu extrahieren (vgl. Kap. 3.2.2),



- Ausreichend repräsentative Testnutzer (vgl. Kap. 4.1.3 u. 4.1.4),
- Standardisiertes Ableiten der Testaufgaben; wenn eine Prüfung aller Arbeitsaufgaben nicht möglich ist, sollten die Primäraufgaben, die wichtigen Sekundäraufgaben und vermutete Probleme aus der Expertenevaluation evaluiert werden (vgl. Kap. 4.1.5),
- Standardisierte Problembewertung (► Kap. 6.3),
- Standardisierte Evaluationsmethoden, die selbst die Testgütekriterien erfüllen (vgl. Kap. 2.4),
- Standardisierter Report (vgl. Kap. 4.1.7),
- Hinweise für den Prüfer auf eventuelle Fehlerquellen und Möglichkeiten, diesen entgegenzuwirken (► Kap. 7),
- Aufspüren und Ausschließen konstanter Störvariablen, welche den Test beeinflussen können. Nicht ausschließbare Störvariablen sollten möglichst konstant gehalten werden.

Um die geforderte Reliabilität für Gebrauchstauglichkeitsuntersuchungen in der Medizin zu verbessern, soll diese Arbeit ein entsprechendes standardisiertes Rahmenwerk bieten.

Zur Überprüfung der Reliabilität muss ein Test nach wiederholter Anwendung bei denselben Personen zu den gleichen Ergebnissen führen. Man unterscheidet Retest-Reliabilität, Paralleltest-Reliabilität und Split-Half-Reliabilität [36, 265]. Bei der Retest-Reliabilität wird der gleiche Test derselben Stichprobe vorgelegt, wobei das Zeitintervall variieren kann; meistens sind es mehrere Wochen. Bei der Paralleltest-Reliabilität wird ein zweiter Test entwickelt, der dieselben Kriterien misst; die Ergebnisse sollten ähnlich ausfallen. Bei der Split-Half-Reliabilität wird an einer Stichprobe ein einzelner Test durchgeführt. Dieser wird in zwei Hälften geteilt, die dann als parallele Tests angesehen werden können. Bei hinreichend großer Ergebnismenge sollten die gemessenen Werte gleich sein.

Reliabilität wird mit der Korrelation zweier Messreihen gemessen, welche von 0 (keine Reliabilität) bis 1 (perfekte Reliabilität) reichen kann. Ein guter Test, der nicht nur zu explorativen Zwecken verwendet wird, sollte eine Reliabilität von über  $r=0,75$  aufweisen [36, 265]. Welche Art der Reliabilitätsprüfung vorgenommen wird, ist hierbei unerheblich. Zusätzlich sollte aufgrund der bereits beschriebenen Problematik bei der Korrelationsmessung die Bland-Altman-Methode verwendet werden [29].

**Validität** gibt an, ob ein Test auch wirklich das misst, was er messen soll [91]. Bei dieser Arbeit muss somit die Frage überprüft werden, ob die Standardprozedur auch die Gebrauchstauglichkeit eines medizinischen Produktes entsprechend der Maße Effektivität, Effizienz und Zufriedenheit evaluiert, und ob diese Ergebnisse in die Realität übertragbar sind.

Validität liegt vor, wenn Ergebnisse eindeutig interpretierbar sind. Dies kann bereits durch eine Konsensbildung erreicht werden. Einigen sich mehrere Personen auf die Glaubwürdigkeit und den Bedeutungsgehalt des Materials, kann dies als ein Hinweis auf Validität angesehen werden [36]. Bei Validität wird unterschieden in:

- **Inhaltsvalidität:** Sie ist gegeben, wenn der Test das zu messende Konstrukt in seinen wichtigsten Punkten erfasst. Im Fall der Standardprozedur ist hierbei die Fragestellung, ob sie Effektivität, Effizienz, Zufriedenheit und Sicherheit erfasst.
- **Kriteriumsvalidität:** Sie ist gegeben, wenn eine Korrelation zwischen Testwerten und Kriteriumswerten einer Stichprobe vorliegt. Fleiß korreliert z.B. mit besseren Noten. Bei der Standardprozedur ist dies beispielsweise der Fall, wenn hohe Effektivität mit besserer Gebrauchstauglichkeit korreliert.
- **Konstruktvalidität:** Sie ist gegeben, wenn aus dem zu messenden Zielkonstrukt Hypothesen ableitbar sind, welche durch die Testwerte bestätigt werden können. Ihr kommt laut Bortz & Döring eine besondere Bedeutung zu, da Inhaltsvalidität kein objektivierbarer Kennwert ist und Kriteriumsvalidität nur bei geeigneten Außenkriterien sinnvoll ist [36].

Ein typisches Validitätsproblem bei der Gebrauchstauglichkeitsuntersuchung ist die Wahl der Versuchspersonen. Da in der Regel nicht alle Nutzer eines Systems in die Untersuchung einbezogen werden können, muss der Test an einer repräsentativen Stichprobe durchgeführt werden. Die Ergebnisse müssen jedoch für die Allgemeinheit gelten. Weitere Validitätsprobleme sind die Verwendung der falschen Arbeitsaufgaben und die Verkennung, dass der Nutzer in seiner normalen Nutzungsumgebung bei der Nutzung bestimmten Einflüssen (wie z.B. Stress, Zeit- oder sozialem Druck) ausgesetzt ist. Weiterhin können Störeffekte wie z.B. Lerneffekte oder der in Kapitel 4.1.3 beschriebene Evaluatoreffekt beeinflussend wirken. Zur Sicherstellung der Validität sind daher folgende Faktoren wichtig:

- Die Verwendung standardisierter, wissenschaftlich validierter Methoden,

- die Vermeidung oder Berücksichtigung von Störeffekten,
- die Auswahl der Stichprobe,
- die Beachtung des Nutzungskontextes,
- die Aufgabenauswahl entsprechend der Aufgabenanalyse.

<b>Thesen:</b>	
T.32	Die Standardprozedur erfüllt die klassischen Testgütekriterien.
T.33	Objektivität wird durch die Standardprozedur erreicht. Die Korrelation zwischen zwei Testanwendern sollte bei einem Wert größer 0,8 liegen.
T.34	Eine hohe Reliabilität kann durch die Standardprozedur erreicht werden. Ein Test nach wiederholter Anwendung führt zu den annähernd gleichen Ergebnissen ( $r=0,75$ ).
T.35	Zumindest bei den schwerwiegendsten Problemen sollte in einem wiederholten Test die gleiche Benotung herauskommen.
T.36	Standardisierte Methoden und eine repräsentative Stichprobe sichern die Validität.

## **4.2 Iterative Entwicklung und Evaluation der „Standardprozedur zur Prüfung der Gebrauchstauglichkeit medizinischer Geräte“**

Für die Entwicklung und Konkretisierung der Standardprozedur wurden, entsprechend dem in Kapitel 3.1 beschriebenen iterativen Entwicklungsprozess, verschiedene Studien durchgeführt. Im Folgenden werden die Evaluationsschritte und die hieraus ermittelten Anforderungen zur Konkretisierung dieser Standardprozedur beschrieben.

### **4.2.1 Entwicklungsstudie I – Studie zur Ermittlung der Gebrauchstauglichkeit von drei Hochfrequenz-Chirurgiegeräten**

#### **Fragestellung:**

In dieser ersten Evaluation sollte der prinzipielle Ablauf einer Gebrauchstauglichkeitsuntersuchung für medizinische Geräte eruiert und die bisherigen theoretischen Überlegungen verifiziert werden. Verschiedene Methoden aus dem Methodenrepertoire des Usability Engineerings (vgl. Kap.2.4.1) wurden ausgewählt und sollten auf ihre Relevanz für eine Gebrauchstauglichkeitsevaluation in der Medizin überprüft werden. Ermittelt werden sollte der genaue Ablauf und die Stichprobengröße der in Kapitel 2.4.2 als erfolgversprechendste Verfahren identifizierten Evaluationsmethoden Heuristische Evaluation und Nutzertest. Weiterhin sollten

die in Kapitel 4.1 ermittelten und in Anhang D zusammengefasst dargestellten Thesen überprüft werden.

### **Material & Methoden:**

Als Evaluationsobjekte wurden drei Hochfrequenzchirurgiegeräte (HF-Geräte) verschiedener Hersteller (Abb. 46) ausgewählt.



**Abb. 46: Hochfrequenzchirurgiegeräte der Gebrauchstauglichkeitsstudie der Firmen KLS Martin, ERBE und BOWA**

Die verwendeten Methoden sind im Einzelnen:

#### **a. Bestimmung des Aufgaben- und Nutzerprofils und Ableiten von Testszenarien**

Zur Ermittlung des Nutzungskontextes (vgl. Kap. 4.1.1) wurden eine Literaturrecherche, eine Expertenbefragung und eine Fragebogenbefragung mit realen Nutzern und Repräsentanten der Hersteller durchgeführt. Teilnehmer dieser Befragung waren die Produktmanager des jeweiligen Herstellers sowie als Vertreter der Nutzergruppe sieben Personen des OP-Personals der Universitätsklinik Tübingen und 29 Auszubildende zum Operations-Technischen Assistenten (OTAs).

#### **b. Heuristische Evaluation**

Als erste Methode der Gebrauchstauglichkeitsbetrachtung wurde eine Heuristische Evaluation durchgeführt. Vier Experten hatten hierbei die Aufgabe, die HF-Geräte zu evaluieren. Zur Evaluation wurden die zehn Heuristiken nach Nielsen (vgl. Kap. 2.4.1.1) und eine sicherheitsabfragende Heuristik verwendet. Die in der Heuristischen Evaluation gefundenen potentiellen Probleme konnten in anschließenden Gebrauchstauglichkeitstests verifiziert werden.

#### **c. Testmaße, -aufgaben und -szenarien**

Aufbauend auf der Heuristischen Evaluation sowie den Nutzerzielen und Arbeitsaufgaben wurden Nutzungsszenarien abgeleitet, welche als Leitfaden für den

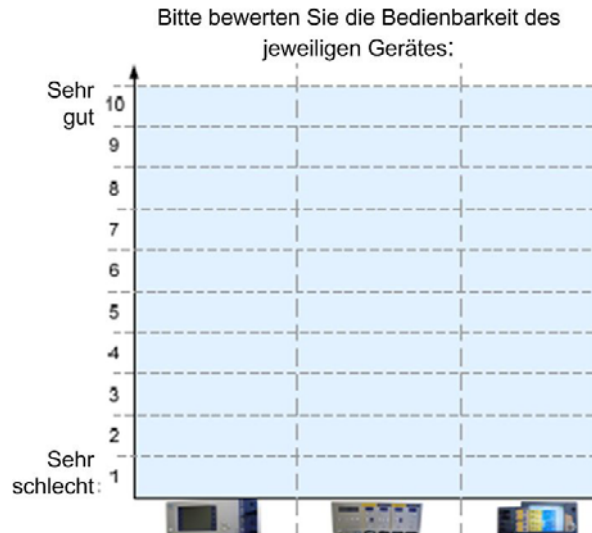
Nutzertest dienen. Hierfür wurden sowohl die gängigen Nutzungsaufgaben als auch einige selten verwendete Aufgaben sowie potentiell problematische Aufgaben, welche in der Heuristischen Evaluation gefunden wurden, verwendet. Die Aufgaben wurden in spezielle Szenarien eingegliedert, um sie innerhalb des Tests realistischer zu gestalten (vgl. Kap. 4.1.5).

Um die Gebrauchstauglichkeit zu erfassen, wurden die Anforderungen an die Maße Effektivität, Effizienz, Zufriedenheit und Sicherheit in messbare Größen überführt und operationalisiert. Hierzu mussten genaue Aussagen definiert und unmissverständlich formuliert werden. Hierauf aufbauend wurden zwölf Testaufgaben (vgl. Anhang E) definiert, um die Gebrauchstauglichkeit der HF-Geräte zu evaluieren. Mit jedem Gerät wurden die gleichen Aufgaben durchgeführt. Um Reihenfolgeeffekte zu vermeiden, wurden die Geräte bei jedem Test in zufälliger Folge getestet.

#### **d. Nutzertest**

Der Nutzertest wurde aufbauend auf den Nutzerzielen und -anforderungen erstellt und entsprechend den bestehenden Standards wie der DIN EN 9241 Teil 11 und 110 und der DIN EN IEC 60601-1-6 im Experimental-OP Tübingen durchgeführt. Hierbei wurden die Testpersonen zuerst in die HF-Geräte eingewiesen, um anschließend vorgegebene Aufgaben bearbeiten zu können. Als Methoden der Datenerhebung wurden eine teilnehmende Beobachtung mit Videodokumentation, Leistungsmessungen (z.B. Aufgabenerledigung, Aufgabenzeiten), die Thinking Aloud-Methode und eine Fragebogenbefragung (vgl. Kap. 2.4.1.2) gewählt. Als weitere Evaluationsmethoden wurden ein aufgabenspezifischer-, ein gerätespezifischer- und ein Vergleichsfragebogen erstellt (vgl. Kap. 4.1.2).

Für den aufgabenspezifischen Fragebogen wurde der „After Scenario-Questionnaire“ (ASQ) (vgl. Kap. 2.4.1.2) angepasst und mit zwei offenen Fragen nach positiven und negativen Einstellungen ergänzt. Dieser Fragebogen wurde nach jeder Aufgabe ausgefüllt. Nach dem Test für ein Gerät füllten die Nutzer einen gerätespezifischen, angepassten ISOMetrics-Fragebogen [17] aus. Nach dem kompletten Test wurde der Vergleichsfragebogen (Abb. 47) ausgefüllt, der die drei Geräte mit Hilfe einer zehnstufigen Skala im direkten Vergleich unter den Kriterien allgemeiner Eindruck, Design, Bedienbarkeit und Zufriedenstellung bewertete.



**Abb. 47: Beispiel aus dem Vergleichsfragebogen zur Ermittlung der subjektiven Zufriedenheit mit der Bedienbarkeit der Testgeräte**

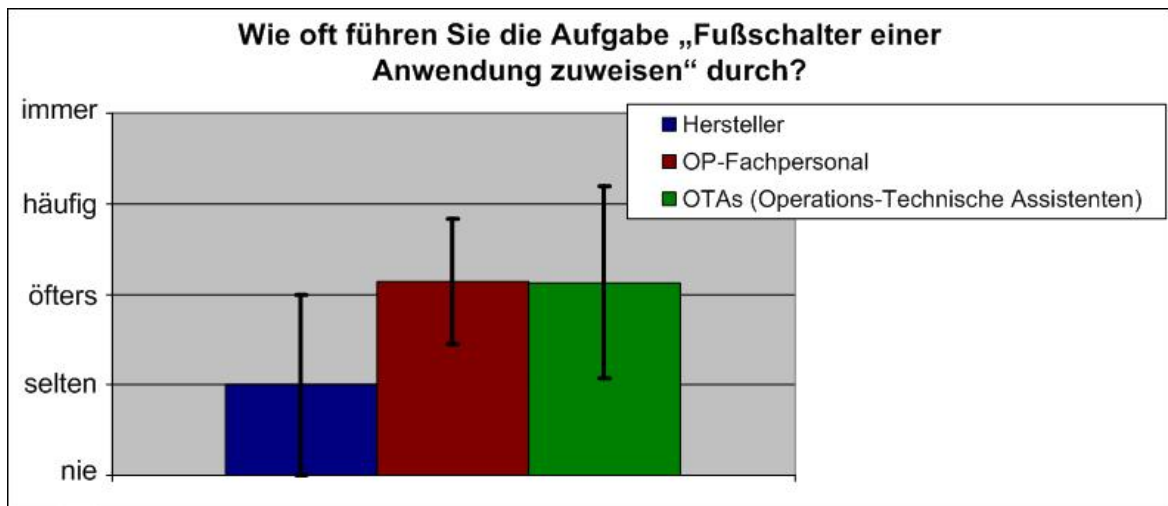
#### **e. Stichprobe:**

Als Stichprobe eignen sich primär die realen Nutzer der HF-Geräte, im Falle dieser Studie sowohl Chirurgen als auch OP-Personal. Dennoch wurde hier von dieser Begrenzung abgewichen, da OP-Personal, welches täglichen Kontakt mit dem Untersuchungsgegenstand hat, mögliche Präferenzen für ein bereits bekanntes Gerät und dessen Bedienkonzept zeigen könnte. Um unvoreingenommene Testnutzer zu erhalten, wurden Probanden mit geringerer Erfahrung bezüglich der zu untersuchenden Systeme ausgewählt, welche jedoch ausbildungstechnisch der Zielgruppe der Geräte entsprachen. Danach bestand die auf 19 Personen angesetzte Stichprobe aus einem Chirurgen, zwei Chirurgieassistenten, welche mit den Geräten seit langem nicht mehr arbeiteten, sowie zwölf Medizinstudenten und vier OTA-Auszubildenden, welche zwar über das nötige Fachwissen zur Bedienung der Geräte verfügten, aber noch nicht für ein bestimmtes Gerät voreingenommen waren. Die Stichprobe unterschied sich weiterhin in Alter und Geschlecht. Der Test wurde sowohl in Testgruppen von zwei Personen als auch mit Einzelnutzern durchgeführt. Vor dem Test erhielten die Teilnehmer eine allgemeine Einweisung in Hochfrequenzchirurgiegeräte.

#### **Ergebnisse und Diskussion:**

Da die Einzelergebnisse dieser Studie für die Methodenfindung der Standardprozedur nur bedingt von Interesse sind wird an dieser Stelle auf deren Einzeldarstel-

lung verzichtet. Deren detaillierte Beschreibung ist der in der Literaturliste angegebenen Publikation und der Diplomarbeit von Baumann zu entnehmen [17, 42]. Die Vorgehensweise, mit der Bestimmung eines Aufgaben- und Nutzerprofils als Grundlage für den Test, mit den hieraus abgeleiteten Testaufgaben und Szenarien, der Auswahl der Stichprobe und der Wahl der Gebrauchstauglichkeitsziele erwies sich als essentiell für die Standardprozedur. Ohne diese Aktionen fehlt das entsprechende Wissen über den Nutzungskontext, um eine Gebrauchstauglichkeitsstudie durchführen zu können. Bei der Aufgabenanalyse zeigte sich, dass Hersteller- und Anwendermeinung teilweise divergieren (Abb. 48). Daher ist es von besonderer Bedeutung, den Anwender in die Analyse des Nutzungskontextes zu integrieren.



**Abb. 48:** Verschiedene Antworten auf die Frage, wie oft eine bestimmte Aufgabe durchgeführt wird veranschaulichen divergierende Ansichten zwischen Herstellern und Anwendern bzgl. der Nutzung der Geräte und ihrer Funktionen.

Nach Reduktion von Mehrfachnennungen konnten insgesamt 65 potentielle Probleme mit der Heuristischen Evaluation gefunden werden. Dies sind mehr Probleme als sich mit dem Nutzertest (39 Probleme) ergaben. 25 Probleme wurden durch beide Methoden aufgedeckt (Tab. 15).

Es wird bestätigt, dass sich durch eine Kombination der beiden Methoden mehr Probleme finden lassen als mit einer Methode allein. Auch stellte sich heraus, dass Auffälligkeiten, welche durch Heuristiken entdeckt wurden, teilweise mehreren Problemen, welche im Nutzertest evaluiert wurden, zugeordnet werden konnten. Dies erklärt auch die im Verhältnis zur Heuristischen Evaluation geringere Gesamtzahl der gefundenen Fehler (Tab. 15). In dieser Studie stellte sich bei der Heuristischen Evaluation als verbesserungswürdig heraus, dass die vorgegebene

nen Heuristiken teilweise doppeldeutig waren, und die Experten nicht genau wussten, welcher Heuristik sie einem Problem zuordnen sollten.

	HE	NT	In HE und NT aufgefallen	Gesamt
HF-Gerät 1	22	14	9	27
HF-Gerät 2	24	11	9	26
HF-Gerät 3	19	14	7	26
$\Sigma$	65	39	25	79

**Tab. 15: Gefundene Probleme aus Heuristischer Evaluation (HE) und Nutzertest (NT), sowie in beiden Methoden aufgefallene Probleme**

Auch zeigte sich, dass einige Auffälligkeiten nicht von allen Experten als problematisch angesehen wurden, und diese in dem Nutzertest nicht auftraten, obwohl sie den Benutzer durch die Wahl der Testaufgaben hätten erreichen müssen. Dies zeigt, dass auch die Expertenmeinung nicht immer repräsentativ für Nutzungsprobleme ist. Aussagekräftig sind nur Probleme, die wirklich beim Nutzer zu beobachten sind. Eine Heuristische Evaluation eignet sich aber besonders, um erste potentielle Probleme zu ermitteln und diese dann in einem anschließenden Nutzertest zu erhärten. Somit könnten neben dem Nutzungskontext und den Gebrauchstauglichkeitszielen auch die Ergebnisse einer Heuristischen Evaluation für die Auswahl der im Test verwendeten Aufgabenszenarien herangezogen werden.

Der Nutzertest zeigte die Probleme auf, welche direkt den Nutzer betreffen. Die Methode „Thinking Aloud“ (vgl. Kap. 2.4.1.2) lieferte viele Daten, welche dem Testleiter sonst verborgen geblieben wären. Teilweise war es jedoch notwendig, zwischen den Aufgaben die Teilnehmer an diese Methode zu erinnern. Hierbei zeigte sich von Vorteil, dass eine teilnehmende Beobachtung (vgl. Kap. 2.4.1.2) durchgeführt wurde, in der die Nutzer direkt angesprochen werden konnten.

Als Methode der Datenerhebung stellte sich als besonders hilfreich der ASQ (vgl. Kap. 2.4.1.2) heraus. Unmittelbar nach der Aufgabendurchführung liefern die Testnutzer damit Daten über ihre Zufriedenheit. Durch die Aufgabenspezifität konnte direkt erkannt werden, bei welcher Aufgabe Nutzer unzufrieden waren. Auch der Vergleichsfragebogen, in welchem Probanden sofort Angaben zur Zufriedenheit machen sollten, erwies sich als leicht verständlich. Er lieferte direkte Vergleichsdaten zwischen verschiedenen Geräten.

Der sehr lange und ausführliche ISOMetrics-Fragebogen (vgl. Kap. 2.4.1.2) zeigte sich in dieser Evaluation als redundant, da er gesamt betrachtet ähnliche Ergebnisse wie die vorher beschriebenen Fragebögen bzgl. einer Zufriedenheitseinstu-



fung lieferte [17]. Er ließ zwar durch seine Mehrdimensionalität leichter die Zufriedenheit speziellen Teilen eines Gerätes zuweisen, jedoch ist dies bei einer rein bewertenden Evaluation nicht notwendig. Daher ist es ratsam, den schnelleren und einfacher auszuwertenden ASQ zu verwenden und in einer abschließenden Befragung allgemeine Fragen zur Zufriedenheit zu stellen.

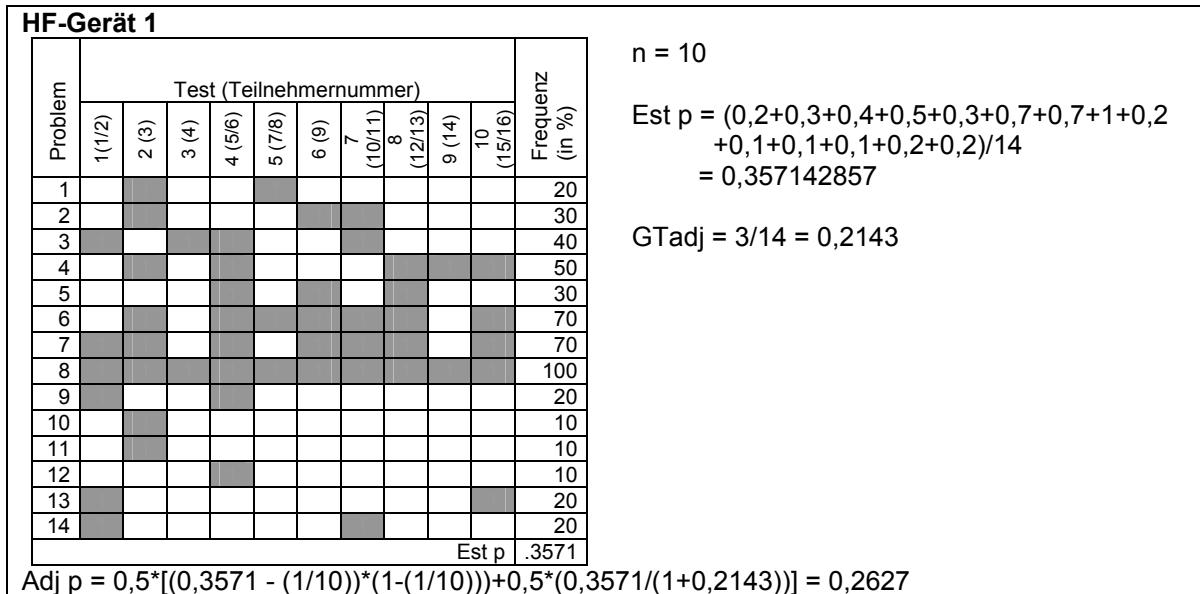
Diese Studie macht hauptsächlich Probleme sichtbar, welche bei gering- bzw. unerfahrenen Nutzern auftreten. Es bleibt offen, inwieweit sie auch bei höherem Erfahrungsgrad vorhanden sind. Daher sollten zukünftig auch erfahrene Nutzer zu einer derartigen Untersuchung eingeladen werden.

Weiterhin sollte in dieser Studie die Anzahl der notwendigen Testteilnehmer herausgefunden werden. Laut Literatur sollten die 16 Teilnehmer, welche sich in zehn Teams gruppierten, ausreichen (vgl. Kap. 4.1.4). Mit Hilfe einer Formel nach Lewis kann ermittelt werden, ob die Stichprobe ausreicht (vgl. Kap. 4.1.4) [241]:

$Adj\ p = \frac{1}{2}[(Est\ p - 1/n)(1 - 1/n)] + \frac{1}{2}[Est\ p / (1 + GTadj)]$	
Adj p =	Angepasster p-Wert
GTadj =	Anzahl der Probleme, die einmal aufgetreten sind, geteilt durch die Gesamtzahl der verschiedenen Probleme
Est p =	Schätzung von p (Prozentzahl der Teilnehmer, die mit dem Problem konfrontiert werden / Anzahl der Probleme)
n =	Stichprobengröße

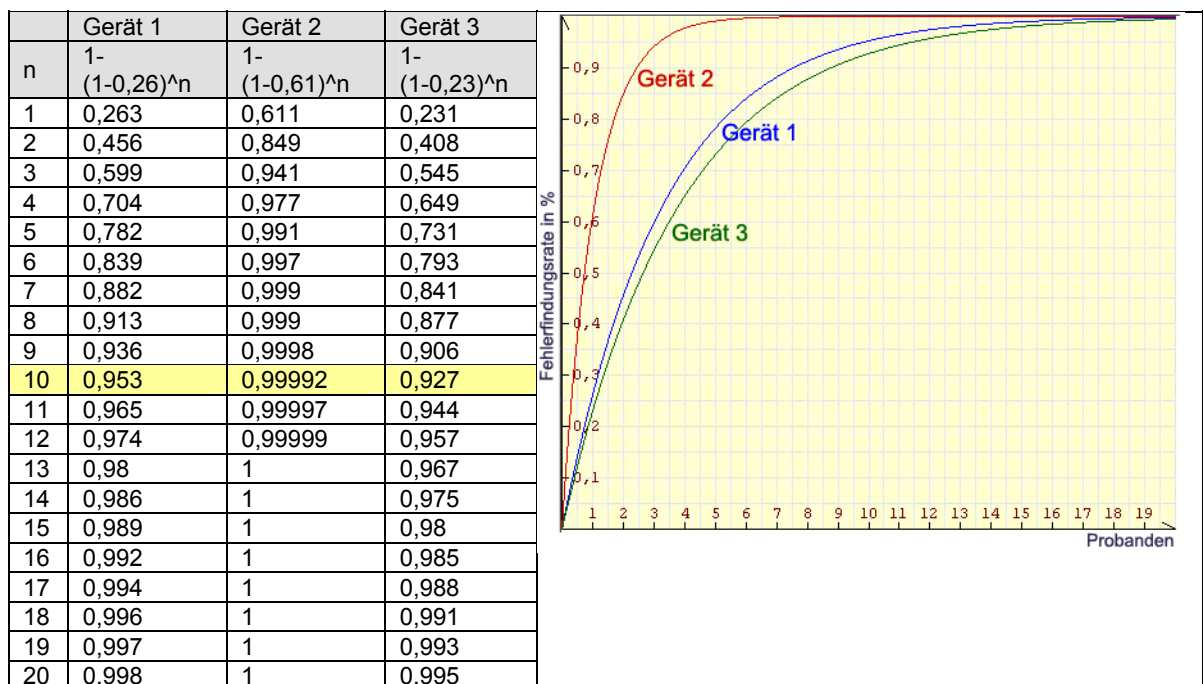
Durch weitere Verwendung der Formel  $(1 - (1 - p)^n)$  kann die Wahrscheinlichkeit für die Entdeckung eines Fehlers berechnet werden.

Bei dem Gebrauchstauglichkeitstest mit zehn Testteams und 16 Teilnehmern wurden bei den ersten beiden HF-Geräten 14 und bei dem dritten HF Gerät 11 Probleme gefunden. Abb. 49 veranschaulicht exemplarisch, bei welchen Testteams welche Probleme auftraten, außerdem die entsprechende Auftretensfrequenz sowie die Berechnung des angepassten p-Werts (Adj p). Weitere Ergebnisse sind im Anhang F dargestellt.



**Abb. 49: Berechnung des angepassten p-Werts (Adj p) für das HF-Gerät 1**

Mit der anfänglich geschätzten Fehlerrate „Est p“ konnte die angepasste Fehlerrate „Adj p“ berechnet werden. Mit einer Stichprobengröße von 10 (Testteams) und Verwendung der Formel  $(1 - (1 - p)^n)$  liegt die Wahrscheinlichkeit für die Entdeckung eines Fehlers (Fehlerfindungsrate) bei Gerät 1 in der Größenordnung von 95,2%, bei Gerät 2 von 99,9% und bei Gerät 3 von 92,7%. Bei einer Fehleranzahl zwischen 11 und 14 heißt dies, dass maximal ein Fehler unentdeckt bleibt. Entsprechende Wahrscheinlichkeiten in Abhängigkeit von der Stichprobe sind in Abb. 50 dargestellt.



**Abb. 50: Berechnung der Fehlerfindungsrate in Abhängigkeit von der Stichprobengröße**

Anhand von Abb. 50 erkennt man, dass zehn Probandengruppen eine ausreichend große Stichprobe darstellen. Die geringste Fehlerfindungsrate ist mit 92,7% ausreichend gut. Für eine Bewertung eines medizinischen Gerätes sollte eine Fehlerfindungsrate um die 90%<sup>40</sup> angesetzt werden. Fraglich ist, ob die Einzelsitzung oder die Verwendung von zwei Probanden zu einem besseren Ergebnis geführt hätte (Tab. 16). Zwar erkennen die Einzelpersonen weniger Probleme als die Teams, jedoch muss hierbei beachtet werden, dass auch der Aufwand bei 12 Teilnehmern in den Teams dreimal so hoch ist wie der Aufwand bei der Einzelsitzung (Tab. 16). Wichtiger ist, darauf zu achten, dass die Stichprobe dem Nutzerprofil entspricht und der Nutzungskontext mit der Stichprobe dargestellt wird. Wird im Nutzungskontext ein Gerät von zwei Anwendern gleichzeitig verwendet, so sollte auch der Gebrauchstauglichkeitstest diese Anwendung simulieren.

	Identifizierte Probleme		Auffälligkeiten insgesamt	
	Einzeltest	Testteam	Einzeltest	Testteam
Gerät 1	10	12	17	33
Gerät 2	10	10	31	45
Gerät 3	8	13	15	31

**Tab. 16: Identifizierte Probleme in einem Einzeltest und einem Test mit zwei Probanden**

Betrachtet man die Ergebnisse der Fehlerfindungsrate, so ist beispielsweise im Falle des Gerätes 2 bei einer Quote von fast 100% bei zehn Nutzern (vgl. Abb. 50) davon auszugehen, dass man alle Fehler des Gerätes mit dem Gebrauchstauglichkeitstest gefunden hat. Die Ergebnisse der Heuristischen Evaluation würden dies jedoch ad absurdum führen, wenn man davon ausgeht, dass nur eins der potentiellen Probleme dieser Expertenevaluation auch ein wirkliches Problem des Gerätes ist. Daher können diese Ergebnisse nur auf die einzelnen Aufgaben bezogen werden, welche getestet werden, und nicht auf das komplette Gerät. Weiterhin ist es wichtig, die Aufgaben des Gebrauchstauglichkeitstests sorgfältig nach den Ergebnissen der Anforderungsanalyse auszuwählen und die in der Heuristischen Evaluation gefundenen Probleme in der Gebrauchstauglichkeitsprüfung zu verifizieren.

Betrachtet man die Unterschiede zwischen Fachpersonal und Laiengruppe, so ist zu erkennen, dass bei den Tests mit dem Fachpersonal (Test 1 und 10) bei allen

<sup>40</sup> In dieser Arbeit, wird nach bisherigen Erkenntnissen (vgl. Kap. 2.1, Kap 4.1.4) und den durchgeführten Studien davon ausgegangen, dass mit einer Problemfindungsrate von 90% ein angemessenes Aufwand-Nutzen-Verhältnis für das medizinische Arbeitssystem gegeben ist. Ein wesentlich höherer Wert würde den Aufwand auf Grund des zu geringen Nutzens nicht rechtfertigen. Ein geringerer Wert ist aufgrund des sicherheitskritischen Aspektes medizinischer Geräte nicht zu vertreten.

Geräten acht Probleme identifiziert werden konnten (vgl. Anhang F). Dies entspricht einer prozentualen Fehlerfindungsrate bei Gerät 1 und Gerät 3 von 57,14% und bei Gerät 2 von 80%. Verglichen mit der Wahrscheinlichkeit für die Entdeckung eines Fehlers in Abhängigkeit von der Stichprobengröße (vgl. Abb. 50) wird erkenntlich, dass das Fachpersonal bei den Geräten 1 (bei  $n=2 \rightarrow 45,64\%$  Wahrscheinlichkeit) und 3 (bei  $n=2 \rightarrow 40,83\%$  Wahrscheinlichkeit) mit mehr Problemen konfrontiert wurde als die Gesamtgruppe. Bei Gerät 2 ist die prozentuale Fehlerfindungsrate fast gleich mit der Wahrscheinlichkeit für die Entdeckung eines Fehlers nach zwei Testpersonen (84,89%).

**Fazit:**

Aus der Methodenbetrachtung dieser Studie können folgende Resultate und Erfahrungen in die Entwicklung der Standardprozedur eingebunden werden:

- Eine einleitende Kontextanalyse ist notwendig, um Anforderungen an die Gebrauchstauglichkeit aufzustellen und Informationen für die Evaluation zu gewinnen. Für diese Kontextanalyse müssen Aussagen der Gerätehersteller und Informationen der Nutzer sowie die eignen Erfahrungen des Prüfers aus einer Feldbeobachtung verwendet werden.
- Heuristische Evaluation und Nutzertest sollten zusammen eingesetzt werden, da sie sich gegenseitig ergänzen und gemeinsam mehr Probleme finden. Die Probleme der Heuristischen Evaluation stellen lediglich potentielle Probleme dar, welche im Nutzertest verifiziert werden müssen. Der Nutzertest zeigt die Probleme auf, welche direkt den Nutzer betreffen.
- Bei der Auswahl der Testaufgaben sind Ergebnisse der Nutzungskontext- und der Anforderungsanalyse sowie der Heuristischen Evaluation zu berücksichtigen.
- Heuristiken nach Nielsen sind für Experten aus der Medizin teilweise doppeldeutig und nicht klar verständlich. Experten haben Schwierigkeiten, Probleme einzelnen Heuristiken zuzuordnen. Daher sollten für das medizinische Arbeitssystem eigene Heuristiken verwendet werden.
- Studenten erwiesen sich als geeignete Probanden, um Probleme aufzudecken. Sie haben die Fähigkeit, Nutzungsnovizen zu repräsentieren, jedoch müssen auch erfahrene Nutzer für die Tests hinzugezogen werden. Bei unerfahrenen Nutzern sind tendenziell mehr Bediendefizite beobachtbar. Mit erfahrenen Nut-

zern kann überprüft werden, wie tief die Probleme in dem Gerät verwurzelt sind. Für eine Bewertung sollten repräsentative Nutzer verschiedener Erfahrungsstufen eingesetzt werden. Um fehlendes Vorwissen bei Novizennutzern auszugleichen, ist eine standardisierte Einweisung ggf. durch den Hersteller notwendig, da schließlich die meisten Geräte auch erst nach einer Einweisung bedient werden dürfen.

- Durch die in Zweiergruppen durchgeführten Tests wurden nicht mehr Probleme gefunden als in Einzelsitzungen. Im Gegenteil werden Probleme, welche bei einem Testnutzer auftreten, durch den zweiten Testnutzer eher kompensiert. Daher sollten Testgruppen nur verwendet werden, wenn wirklich zwei Personen zur Interaktion mit dem Gerät notwendig sind.
- Die Anzahl der Probanden hängt vom Nutzungskontext ab. Werden in der Kontextbetrachtung verschiedene Nutzergruppen identifiziert, dann muss die Stichprobe entsprechend höher ausfallen. Ist die Nutzergruppe homogen, kann die Stichprobegröße kleiner sein. In dieser Studie zeigten sich neun Testteams als ausreichend. Mit dieser Anzahl konnte eine Fehlerfindungsrate von über 90% für alle Geräte erreicht werden.
- Eindeutige Größen für Effektivität und Effizienz müssen bestimmt und gemessen werden. Es reicht nicht aus, die Zeit zur Aufgabenerfüllung zu messen, es müssen auch die einzelnen Arbeitsschritte (Subtasks) zeitlich dokumentiert werden. Auch beschäftigt sich der Anwender nicht immer ununterbrochen mit der gesamten Aufgabe. Es hat sich in dieser Studie bewährt, genaue Start- und Endzeitpunkte der Messung festzulegen, um Leerläufe aus der Berechnung der Effizienz zu eliminieren.
- Die Methode „Thinking Aloud“ liefert Daten, welche durch Leistungsmessungen oder Fragebögen nicht erhoben werden können.
- Der ASQ eignet sich zur aufgabenspezifischen Zufriedenheitsmessung. Es ist sinnvoll, beim ASQ zusätzlich eine Frage über das subjektive Empfinden der Bearbeitungszeit zu stellen. Diese Ergebnisse können in die Bewertung der Aufgabenzeit einfließen.
- Vergleichsfragebogen und IsoMetrics liefern vergleichbare Daten. Die Methode ist hierbei nicht entscheidend. Wichtiger ist, dass überhaupt ein abschließender Fragebogen zur Bestimmung der subjektiven Zufriedenheit durchgeführt wird. Der Vergleichsfragebogen ist für die Nutzer einfacher zu verwenden.
- Durch die verwendeten Methoden können Gebrauchstauglichkeitsprobleme

identifiziert und Daten über Effektivität, Effizienz und Nutzerzufriedenheit erhoben werden. Die Bewertung der Gebrauchstauglichkeit ist mit ihnen jedoch nicht möglich. Hierfür ist eine weitere Methode erforderlich.

#### **4.2.2 Entwicklungsstudie II – Studie zur Ermittlung der Gebrauchstauglichkeit von MIC Handgriffen**

##### **Fragestellung:**

Ausgehend von den Ergebnissen der ersten Studie, wurde zur Methodenkonkretisierung in einer zweiten Studie die Gebrauchstauglichkeit von minimalinvasiven Handgriffen überprüft. Die in der ersten Studie als geeignet befundenen Ansätze sollten weiter erprobt und durch neue Erkenntnisse ergänzt werden. Bei den evaluierten chirurgischen Instrumenten handelt es sich um passive Medizinprodukte (vgl. Kap. 1.2.2.3). Es wurde mit dieser Studie überprüft, ob die bisher entsprechend der DIN EN IEC 60601-1-6 und der DIN EN 9241-11 entwickelte Methode auch den neuen Anforderungen der 2008 harmonisierten Norm DIN EN IEC 62366 [104] gerecht wird.

Die Studie wurde im Experimental-OP des Universitätsklinikums Tübingen und am St. Olavs Universitätskrankenhaus Trondheim (Norwegen) durchgeführt. Durch diese beiden unterschiedlichen Untersuchungsorte sollte überprüft werden, ob der Ort der Bewertung oder nationale Präferenzen auf die Gebrauchstauglichkeitsbewertung einen Einfluss haben.

Weiterhin sollten im Rahmen dieser Studie der Unterschied zwischen Novizen und erfahrenen Nutzern und die Verwendung von speziell für den Untersuchungsgegenstand entwickelten Heuristiken in der Expertenevaluation überprüft werden.

Für die Ergebnisdokumentation wurde das Industrieformat für Berichte über Gebrauchstauglichkeitsprüfungen entsprechend der Norm ISO 25062 auf seine Anwendbarkeit in der Medizin überprüft [200].

##### **Material und Methoden:**

Für die Studie wurden Einmalhandgriffe mit der Funktion einer Präparier- und Faszange sowie einer Schere von vier verschiedenen Herstellern ausgewählt und hinsichtlich Gebrauchstauglichkeit und Ergonomie untersucht. Bei dieser Studie wurde primär der ErgoGrip®, ein von der norwegischen Firma Surgitech

entwickelter Prototyp, betrachtet. Die drei weiteren Handgriffe (im Folgenden als Griff 1 bis 3 bezeichnet) dienen zur Ermittlung von Vergleichsdaten. Die verwendeten Methoden werden im Folgenden aufgelistet:

### a. Bestimmung des Aufgaben- und Nutzerprofils und Ableitung von Test-szenarien

Nach den positiven Erfahrungen der ersten Studie wurden auch in dieser Studie zu Beginn der Nutzungskontext ermittelt und hieraus die Nutzerziele abgeleitet. Die Daten wurden hierbei aus der Literatur, aus Expertenbefragungen und einer Feldbeobachtung generiert. Die Ergebnisse der Aufgabenanalyse wurden in Form einer HTA (vgl. Kap. 4.1.1.2) dokumentiert (Abb. 51).

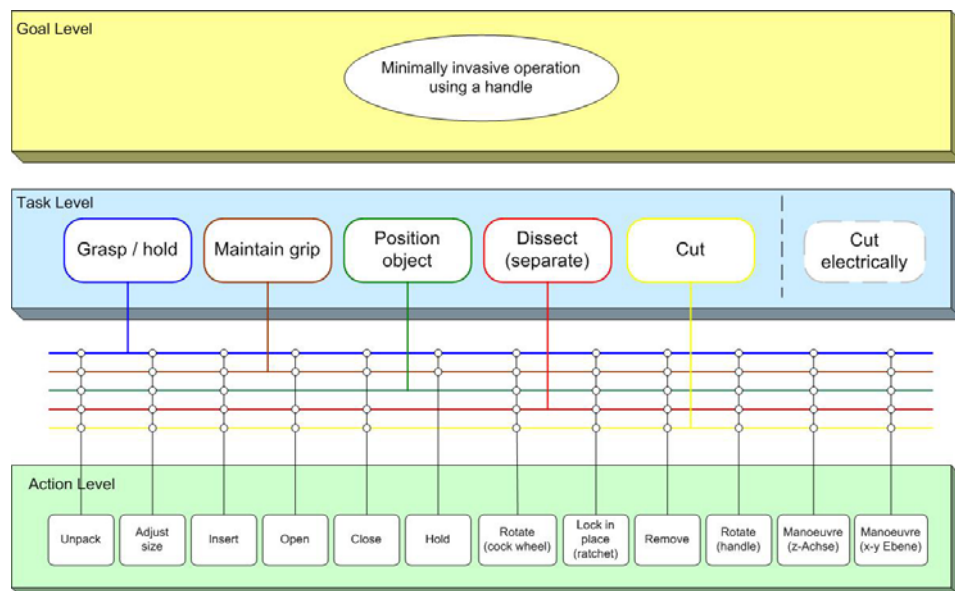


Abb. 51: Hierarchische Aufgabenanalyse für den ErgoGrip®

### b. Heuristische Evaluation

Auch bei dieser Gebrauchstauglichkeitsprüfung wurde die bewährte Expertenevaluation durchgeführt. In der vorherigen Studie hatte sich gezeigt, dass die verwendeten Heuristiken aufgrund ihrer Softwareausrichtung teilweise missverständlich waren. Daher wurden nun speziell für laparoskopische Handgriffe entwickelte Heuristiken verwendet [255], um potentielle Schwachstellen zu finden, welche dann in den folgenden Nutzertests verifiziert werden sollten.

### c. Testmaße und -szenarien

Testmaße und -szenarien wurden entsprechend der in der ersten Studie bewährten Methode entwickelt. In dieser Studie stellte sich in der Nutzungskontextanaly-

se heraus, dass die ergonomische Handhabung eine besondere Anforderung an die Handgriffe darstellt. Zur Evaluation dieser Anforderung mussten spezielle Testaufgaben definiert werden. Hierunter fielen auch Beanspruchungsmessungen mittels EMG<sup>41</sup>, Beobachtung von Druckstellen sowie die Messung der subjektiven Zufriedenheit bei der Handhabung.

#### **d. Nutzertest**

Der Nutzertest bestand aus insgesamt acht Aufgaben, von denen fünf typische Arbeitsaufgaben den Nutzungskontext repräsentierten, und drei besonders spezielle Aktionen prüften. Der Testablauf gliederte sich weitgehend wie der Ablauf der ersten Studie. Verwendet wurden wiederum die Methoden: Teilnehmende Beobachtung, „Thinking Aloud“ und Videoanalyse. Als Fragebogen wurde nach jeder Aufgabe der bereits bewährte ASQ verwendet. Zusätzlich wurden Fragen der subjektiven Einschätzung zu empfundenen Schmerzen, Druckstellen und Verspannungen gestellt (vgl. Anhang G). Nach dem Test eines Handgriffs füllten die Testnutzer einen handgriffspezifischen Fragebogen aus, welcher spezielle Zufriedenheitseinschätzungen zu den einzelnen Arbeitsaufgaben dokumentierte, sowie ein Semantisches Differential [36]. Nach dem kompletten Test wurden wiederum ein vergleichender Fragebogen und ein persönlicher Fragebogen zur Ermittlung der demografischen Daten der Testnutzer eingesetzt. Zusätzlich erfolgte ein Ranking, bei welchem die Testnutzer die Geräte entsprechend ihrer Zufriedenheit platzieren sollten.

#### **e. Stichprobe**

Die Teilnehmer dieser Studie wurden entsprechend dem Nutzerprofil an den Unikliniken in Trondheim und Tübingen rekrutiert. Für diese Studie wurden Testnutzer herangezogen, welche sich sowohl in unterschiedliche Repräsentanten der tatsächlichen als auch der potentiellen Nutzergruppe (Novizen / Studenten) gliederte. Des Weiteren wurde der Test mit Ergonomieexperten durchgeführt. Die Probanden repräsentierten unterschiedliche Altersgruppen, Größen, Geschlechter, Links- und Rechtshänder. Insgesamt wurden für den Nutzertest 25 Teilnehmer ausgewählt: 13 Chirurgen und 12 Novizen.

---

<sup>41</sup> EMG = Elektromyographie: elektrophysiologische Messung, bei der die elektrische Muskelaktivität gemessen wird.



**f. Testumgebung**

Der Nutzertest wurde im Experimental-OP in Tübingen und am Universitätsklinikum Trondheim durchgeführt. Die Testaufgaben wurden an einem Pelvitainer absolviert. Wegen der Durchführung des Tests an zwei unterschiedlichen Standorten gab es geringe Unterschiede im Testsetup.

**g. Testreport**

Im Sinne einer standardisierten Testberichterstattung wurde ein den Anforderungen des internationalen Standards ISO 25062 [200] für „Gebrauchstauglichkeitsberichte für Softwareprodukte“ entsprechender Report für Medizinprodukte angepasst (Kap. 4.1.7) und verwendet.

**Ergebnisse und Diskussion**

Wie bereits bei der vorherigen Studie werden auch bei dieser Studie in dieser Arbeit die Ergebnisse präsentiert, welche für die Entwicklung der Standardprozedur relevant sind. Weitere Ergebnisse sind der aus dieser Studie entstandenen Publikation zu entnehmen [43].

Auch in diesem Test bewährte sich wiederum die anfängliche Erhebung des Nutzungskontextes. Es zeigte sich, dass nicht immer intensive Befragungen wie in der ersten Studie durchgeführt werden müssen, sondern in vielen Fällen Literaturrecherche, Expertengespräch und Überprüfung dieser Ergebnisse durch eine Kontextbegehung ausreichen. Die Dokumentation mittels der HTA erwies sich als sinnvoll, da sie auf einfache und verständliche Weise das Ziel- und Aufgabengefüge der Nutzer veranschaulicht.

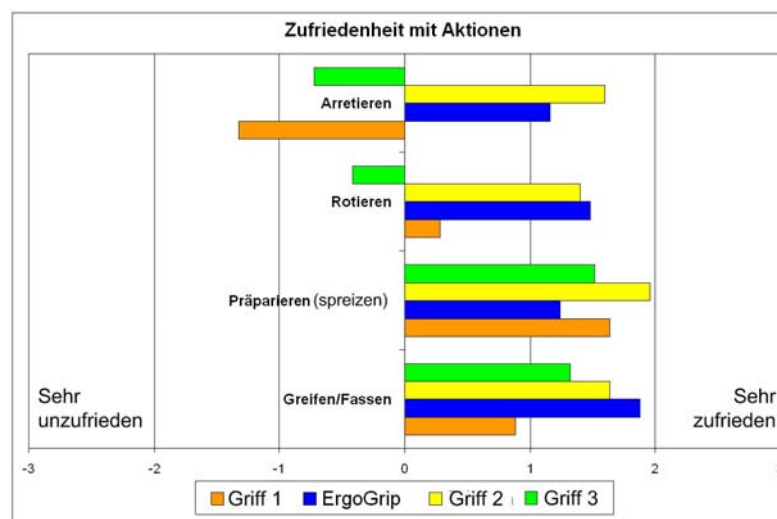
Die Kombination der Methoden Heuristische Evaluation und Nutzertest hat sich wieder bewährt. Im Gegensatz zur ersten Studie zeigte sich jedoch, dass hier die meisten Probleme mit dem Nutzertest gefunden wurden (Tab. 17). Lediglich eins der insgesamt acht gefundenen Probleme wurde allein durch die Heuristische Evaluation gefunden, weitere vier durch beide Methoden und drei nur durch den Nutzertest. Dies lag u.a. daran, dass die Testaufgaben teilweise extra zur Verifikation dieser Probleme ausgerichtet waren.

Als sehr sinnvoll erwiesen sich die speziell angepassten Heuristiken für die Expertenevaluation. Diese waren nicht nur spezifisch auf das zu untersuchende Medizinprodukt zugeschnitten, sondern auch in einem für die Experten vertrauten Wortlaut formuliert.

Problem	NT	HE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	$\Sigma$	Frequenz
1	X																										6	24,00	
2	X																										9	36,00	
3	X	X																									10	40,00	
4	X	X																									4	16,00	
5	X	X																									4	16,00	
6	X																										7	28,00	
7	X	X																									6	24,00	
8		X																									0		
																										26,29			

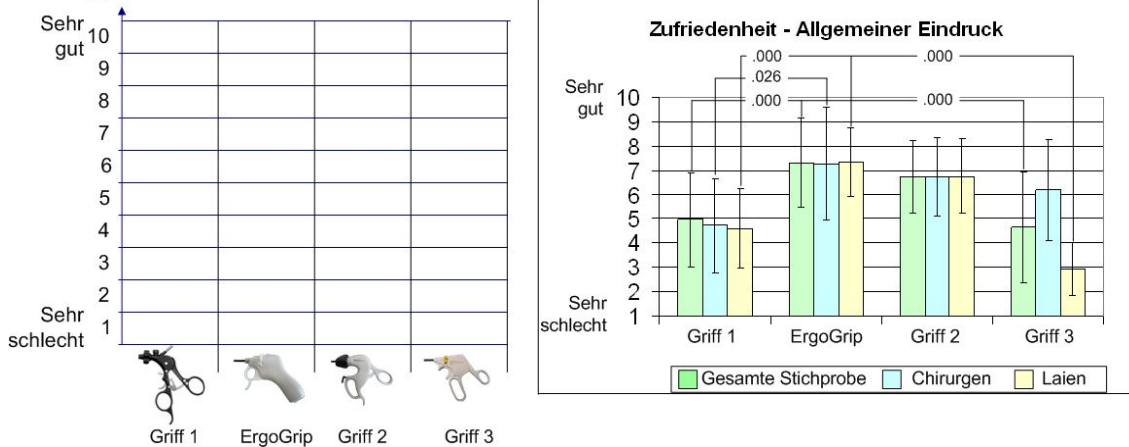
**Tab. 17: Gefundene Probleme bei einem Handgriff durch Heuristische Evaluation (HE) und Nutzertest (NT)**

Bei dem Nutzertest zeigte sich neben den bereits bewährten Methoden, dass der handgriffspezifische Fragebogen aussagekräftige Ergebnisse liefert (vgl. Anhang H). Durch die Ermittlung der subjektiven Zufriedenheit zu einzelnen Aktionen und Arbeitsaufgaben (Abb. 52) konnten diese speziell bewertet werden. Als eher ungeeignet stellte sich das verwendete Semantische Differential heraus, da die Nutzer diese Art der Befragung teilweise nicht verstanden, und diese Art der Datenerhebung nur bedingt nutzbare Ergebnisse für eine Bewertung lieferte. Bei Vergleichstests ist es sinnvoll, am Ende ein Ranking bzgl. der Fragestellung durchzuführen, welches Gerät der Nutzer favorisiert (Abb. 53). Das Ergebnis dieses Rankings ist jedoch nicht überzubewerten, da neben der subjektiven Nutzerzufriedenheit auch noch die Maße Effektivität, Effizienz und Sicherheit für die Gebrauchstauglichkeit von medizinischen Geräten relevant sind.



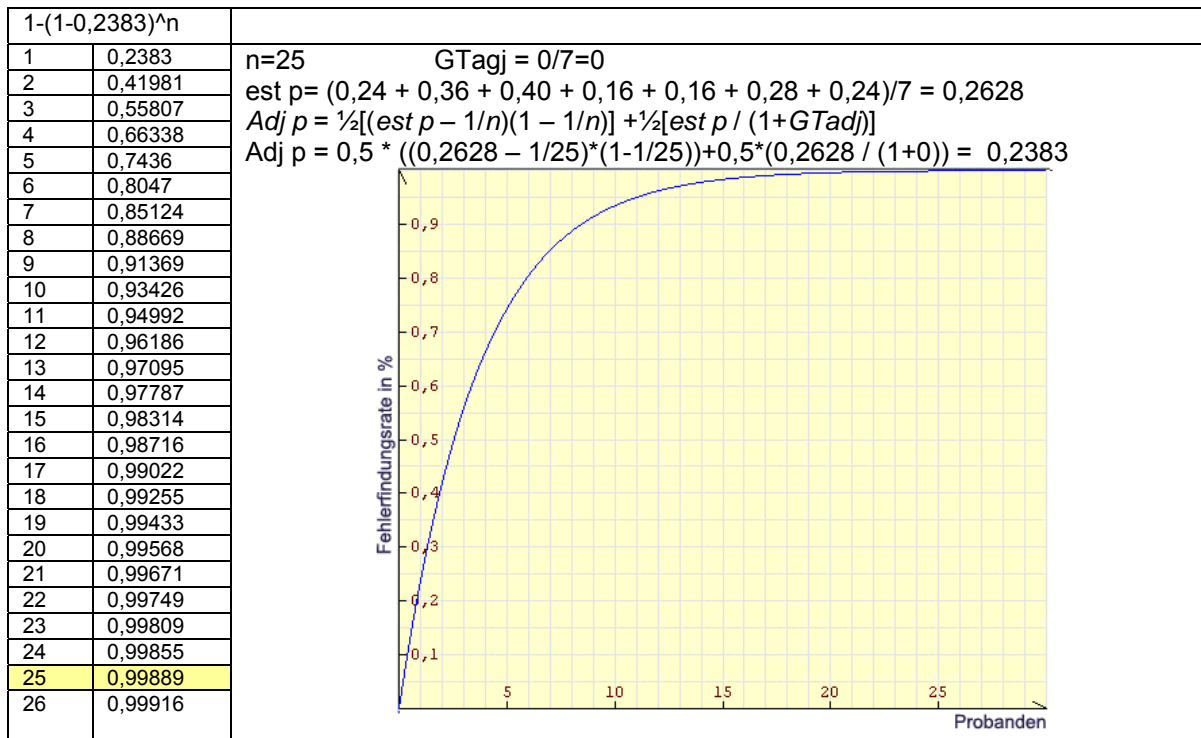
**Abb. 52: Beispiele für die Zufriedenheit mit Aktionen von minimalinvasiven Handgriffen**

Beurteilen Sie die Handgriffe im Hinblick auf ihren allgemeinen Eindruck.



**Abb. 53: Fragebogen und Ergebnisse des vergleichenden Fragebogens bezüglich des allgemeinen Eindrucks, signifikante Unterschiede ( $p < 0,05$ ) sind eingezeichnet**

In der Anforderungsanalyse stellte sich heraus, dass die Nutzung der Handgriffe mit speziellen Anforderungen (Stellung und Haltung des Anwenders, Greifart, Bewegung, Kraftaufwand etc.) verbunden ist, und somit unterschiedliche körperliche Belastungen oder Beanspruchungen bei der Bedienung auftreten können. Um dies zu messen konnte die Elektromyographie sinnvoll eingesetzt werden. Druckstellen konnten durch Beobachtung und subjektive Einschätzung gemessen werden. Schmerzen bei der Bedienung wurden subjektiv durch Befragung ermittelt.



**Abb. 54: Berechnung der Fehlerfindungsrate in Abhängigkeit von der Stichprobengröße**

Die Betrachtung der Stichprobengröße und die Berechnung der Fehlerfindungsrate, entsprechend der in Studie 1 beschriebenen Methode (Abb. 54), zeigen, dass mit der verwendeten Stichprobengröße von 25 Teilnehmern eine Problemfindungsrate von über 99% erreicht werden konnte. In dieser Studie hätte bereits eine Stichprobengröße von neun Teilnehmern genügt, um eine Problemfindungsrate von über 90% zu erreichen.

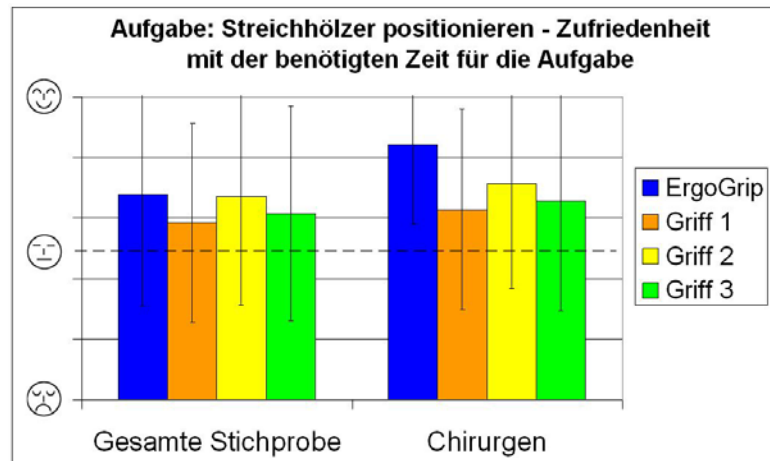
Innerhalb der Stichprobe konnte zwischen Expertennutzern (Chirurgen) und Laien unterschieden werden. Es war zu erkennen, dass die Ergebnisse für Effektivität, Effizienz und Zufriedenheit bei den verschiedenen Nutzergruppen signifikant abwichen. Dennoch waren die Abweichungen zwischen den Ergebnissen in der Regel vergleichbar. Beispielsweise erhielten Handgriffe, welche bei den Chirurgen gut abschnitten, auch entsprechend gute Werte bei den Laien. Verständlicherweise benötigten die Laien bei einigen Aufgaben länger als die erfahrenen Chirurgen, jedoch waren die zeitlichen Unterschiede vergleichbar (Tab. 18). Entscheidend zur Beurteilung der Handgriffe war jedoch, dass die benötigten Zeiten beider Nutzergruppen für die jeweiligen Handgriffe signifikant unterschiedlich waren. Die Handgriffe mit denen die Aufgaben am schnellsten bzw. am langsamsten bewältigt werden konnten waren bei beiden Nutzergruppen die gleichen. Auch in Bezug auf Effizienz und Zufriedenheit waren die Ergebnisse ähnlich. Tab. 19 und Abb. 55 veranschaulichen dies anhand der gemachten Fehler und der Zufriedenheit der Nutzer.

	ErgoGrip	Griff 1	Griff 2	Griff 3	Ø aller Handgriffe
Ø-Aufgabenzeit in min:sek	01:57	02:25	02:34	02:46	02:26 Toleranz: ±15s
Abweichung vom Ø aller Handgriffe	-00:29	+00:01	+00:08	+00:20	
Ø-Aufgabenzeit in min:sek (Chirurgen)	1:20	1:23	1:47	1:48	1:34 Toleranz: ±9s
Abweichung vom Ø aller Handgriffe	-00:14	-00:11	+00:13	+00:14	
Ø-Aufgabenzeit in min:sek (Laien)	2:37	3:32	3:25	3:49	3:21 Toleranz: ±20s
Abweichung vom Ø aller Handgriffe	-00:44	+00:11	+00:05	+00:28	

Tab. 18: Beispiel für die Zeitmessung einer Aufgabe

	ErgoGrip	Griff 1	Griff 2	Griff 3
∑ Fehler: Alle Nutzer (n=25)	7	7	10	10
∑ Fehler: Chirurgen (n=13)	3	3	4	4
∑ Fehler: Laien (n=12)	4	4	6	6
Fehler = Verletzung der Gallenblase				

Tab. 19: Beispiele für die Fehlermessung



**Abb. 55: Beispiel für die Auswertung des ASQs „Zufriedenheit mit der Aufgabe“**

In dieser Studie sind gewisse Abweichungen in den Antworten der Fragebögen zwischen den Nutzergruppen zu erkennen, was durch unterschiedliche Sichtweisen und Einstellungen der Nutzer erklärt werden kann. Daher müssen unbedingt alle identifizierten Nutzergruppen für eine Gebrauchstauglichkeitsprüfung herangezogen werden. Für die Bewertung einzelner Aufgaben ist die genaue Betrachtung der einzelnen Variablen, Maße und Messkriterien von Bedeutung. Dies ist umso wichtiger, je komplexer die zu prüfenden Medizinprodukte sind und je mehr Fachwissen sie von ihren Anwendern voraussetzen. Die in dieser Studie getesteten Handgriffe differenzieren sich primär durch eine angenehme und ergonomische Verwendung und eine ausreichende Funktionalität. Ergonomische Defizite, welche Auswirkungen auf die Handhabung nach sich ziehen, betreffen meist alle Anwender, soweit sie das Gerät richtig verwenden.

Weiterhin hat sich der für Medizinprodukte angepasste Report nach ISO 25062 (vgl. Kap. 4.1.7) bewährt.

Insgesamt erweist sich die für aktive Medizinprodukte entsprechend der DIN EN IEC 60601-1-6 entwickelte Standardprozedur in ihrem Grundrahmen auch für passive Medizinprodukte im Sinne der DIN EN IEC 62366 als anwendbar.

#### **Fazit:**

Folgende Studienergebnisse fließen in die Weiterentwicklung der Standardprozedur ein:

- Die Hierarchische Aufgabenanalyse (HTA) (Kap. 4.1.1.2) veranschaulicht auf einfache und verständliche Weise das Ziel- und Aufgabengefüge der Nutzer. Sie ist eine geeignete Methode zur Bestimmung des Nutzungskontextes.

- Bei Geräten, welche spezielle Anforderungen an Anatomie und Biomechanik haben, sollten Belastungsmessungen (z.B. Elektromyographie oder Kraftmessung) durchgeführt werden.
- Auch passive medizinische Geräte können mit der in dieser Arbeit entwickelten Methode evaluiert werden.
- Speziell an medizinische Produkte angepasste Heuristiken erweisen sich als geeigneter als die allgemein bestehenden Heuristiken interaktiver Geräte.
- Ein gerätespezifischer Fragebogen zu einzelnen Nutzeraktionen eignet sich besonders bei Geräten, mit denen Benutzer offensichtliche Aktionen durchführen, welche mit einem speziellen Bauteil oder Element des Gerätes verbunden sind, z.B. bei der Verwendung der Arretierungsfunktion an den getesteten Handgriffen oder bei der Verwendung einer Hilfe im Falle eines interaktiven Produktes.
- Durch den Einsatz eines standardisierten Reports wird ein Mindestmaß an Dokumentation festgelegt. Hierdurch werden bestimmte Teile eines Tests obligatorisch, was auch zur Reliabilität des Tests sowie zur Vergleichbarkeit der Ergebnisse mit denen anderer Tests beiträgt.
- In dieser Studie zeigt sich, dass eine Stichprobengröße von neun Teilnehmern ausreichend gewesen wäre, um eine Problemfindungsrate von über 90% zu erreichen.

### 4.2.3 Entwicklungsstudie III – Evaluation einer Spritzenpumpe

#### **Fragestellung:**

Ausgehend von den Ergebnissen der bisherigen beiden Studien, wurde zur Methodenkonkretisierung und Evaluation der Standardprozedur in einer dritten Studie die Gebrauchstauglichkeit einer Spritzenpumpe überprüft. Die als geeignet befundenen Ansätze wurden weiter erprobt und durch neue Erkenntnisse ergänzt. Hierbei sollte zum ersten Mal überprüft werden, ob auch Dritte die Standardprozedur anwenden können.

#### **Material und Methoden:**

Für die Studie wurde eine Spritzenpumpe (Abb. 56) ausgewählt und hinsichtlich ihrer Gebrauchstauglichkeit evaluiert. Dieses Medizinprodukt dient der kontinuierlichen dosierten, intravenösen Medikation eines Patienten. Über einen bestimm-

ten Zeitraum verabreicht es dem Patienten aus einer Spritze über eine Leitung selbstständig eine programmierte Medikamentendosis.



**Abb. 56: Untersuchungsgegenstand der Spritzenpumpenstudie**

Die für die Gebrauchstauglichkeitsevaluation verwendeten Methoden werden im Folgenden aufgelistet:

#### **a. Bestimmung des Nutzungskontextes und Ableiten von Testszenarien**

Nach bewährtem Vorbild aus den vorherigen Studien, wurden in dieser Studie zu Beginn der Nutzungskontext ermittelt und hieraus die Nutzerziele abgeleitet. Zum Einsatz kamen Literaturrecherche, eine Feldbeobachtung, ein Herstellergespräch und eine Nutzer- und Expertenbefragung. Daraus wurden die Testszenarien abgeleitet.

#### **b. Expertenbegutachtung**

Um neben der erprobten Heuristischen Evaluation weitere Methoden auf ihre Relevanz für die Standardprozedur zu evaluieren, wurden ersatzweise eine Expertenevaluation durch den Testleiter und zwei Expertengespräche durchgeführt. Als Experten dienten ein erfahrener Nutzer (Anästhesiepfleger des Universitätsklinikums Tübingen) und ein auf Gebrauchsanweisungen spezialisierter technischer Redakteur für die Begleitdokumentation.

#### **c. Testmaße und -szenarien**

Ausgehend von der Analyse des Nutzungskontexts und den Ergebnissen der Experteninspektion, wurden die Anforderungen an das Produkt und Gebrauchstauglichkeitsziele abgeleitet. Zur Definition der Testaufgaben wurden Ergebnisse

der Nutzungskontextanalyse, speziell der Aufgabenanalyse und der Expertenevaluation herangezogen. Für den Test wurden alle Primäraufgaben, einige Sekundäraufgaben und potentiell problembehaftete Aufgaben ausgewählt.

#### **d. Nutzertest**

Der Nutzertest bestand aus insgesamt 19 Aufgaben (vgl. Anhang I). Der Testablauf gliederte sich weitgehend wie der Ablauf der ersten Studien. Wiederum wurden die Methoden Teilnehmende Beobachtung, „Thinking Aloud“ und Videoanalyse angewandt. Nach jeder Aufgabe kam der bereits bewährte ASQ zur Anwendung. Zur Ermittlung der Zufriedenheit, mit dem Gerät, wurde im Anschluss an die Aufgaben ein neuer Fragebogen eingesetzt, welcher Teile des IsoMetrics und des PSSUQ (vgl. Kap. 2.4.1.2) miteinander kombinierte. Die Fragen wurden entsprechend dem medizinischen Kontext angepasst. Weiterhin wurde ein persönlicher Fragebogen zur Ermittlung der demografischen Daten der Testnutzer verwendet. Als Evaluationsleiter fungierte in diesem Test ein Masterstudent des Studiengangs „Communication Planning and Design“ der Hochschule für Gestaltung Schwäbisch Gmünd.

#### **e. Stichprobe**

In dieser Studie wurde, entsprechend den bisherigen Erfahrungen, die Stichprobe auf neun Testpersonen angesetzt. Kriterium für die Teilnehmer war, dass sie dem in der Nutzungskontextanalyse ermittelten Benutzerprofil entsprachen und eine nach §5 MPBetreibV geforderte Einweisung in das medizinische Gerät durch den Hersteller erhalten hatten. An dem Test nahmen zwei Krankenpfleger, ein Anästhesiepfleger, fünf Auszubildende zum Anästhesie-Technischen-Assistenten (ATA) und ein Krankenhaushygieniker teil.

#### **f. Testumgebung**

Der Nutzertest wurde im Experimental-OP Tübingen durchgeführt. Er fand in zwei unterschiedlichen Beleuchtungssituationen statt, um zu überprüfen, inwieweit auch die Lichtverhältnisse die Gebrauchstauglichkeit der Spritzenpumpe beeinflussen. Abb. 57 veranschaulicht den Testaufbau.





**Abb. 57: Testumgebung:** Die Ziffern 1-4 zeigen die Kamerapositionen und -ausrichtung, der rote Punkt markiert die Position der Testperson, der graue Punkt die des Testleiters, das graue Rechteck die des Testobjektes. Weiterhin zu sehen sind der OP-Tisch in der Mitte des Raumes sowie die Arbeitstische mit dem Testmaterial.

### g. Testreport

In dieser Studie wurde kein spezieller Report angefertigt. Die Ergebnisse wurden innerhalb der Masterarbeit [357] dokumentiert, welche die Komponenten des festgelegten Reports enthielt (vgl. Kap. 4.1.7).

### Ergebnisse und Diskussion:

Die Detailergebnisse dieser Studie finden sich in der Masterarbeit von Schiendzielorz [357] und der bisher unveröffentlichten Publikation [45]. Es werden im Folgenden nur Ergebnisse vorgestellt, die für die Weiterentwicklung der Standardprozedur relevant sind.

Auch in diesem Test haben sich wiederum die anfängliche Erhebung des Nutzungskontextes, die hieraus abgeleitete Anforderung an das Produkt und die ermittelten Gebrauchstauglichkeitsziele bewährt. Auch die Kombination einer Experteninspektion mit einem Nutzertest zeigte sich erneut als sinnvoll. Es zeigte sich allerdings in dieser Studie, dass im Vergleich zu den bisher verwendeten Expertenevaluationen, in denen Heuristiken eingesetzt wurden, deutlich weniger potentielle Probleme evaluiert werden konnten (Tab. 20). Somit scheint in der Medizin bei einer Expertenevaluation ein strukturiertes Vorgehen mit Hilfe von

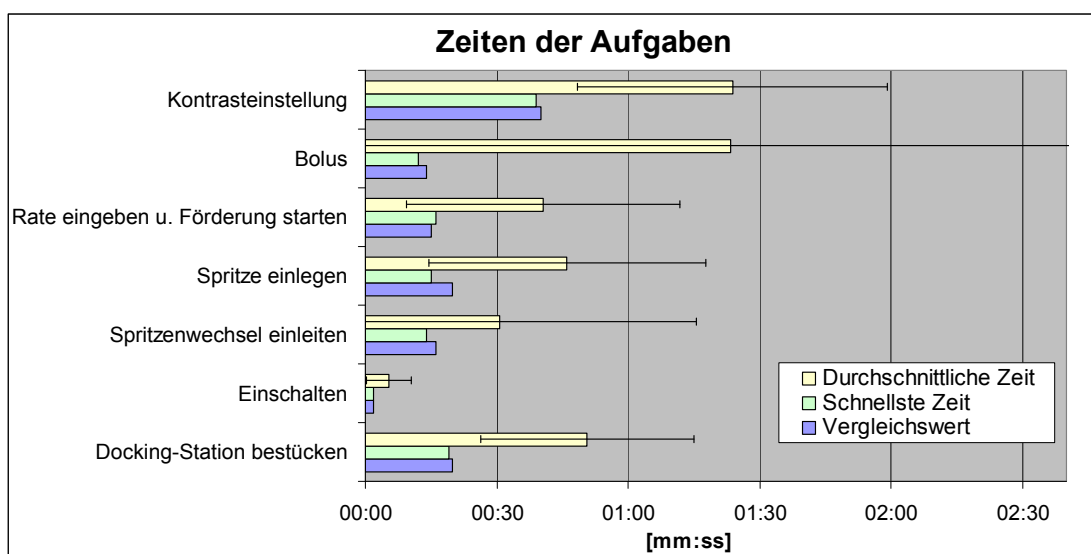
Heuristiken, wie für andere Bereiche durch verschiedene Studien (vgl. Kap. 2.4.2) belegt, effizienter zu sein als eine unstrukturierte Befragung.

Studie	Testgerät	Im Nutzertest gefundene Probleme	davon in Experten-evaluation aufgefallen	%
1	HF-Gerät 1	14	9	64,29
	HF-Gerät 2	11	10	90,91
	HF-Gerät 3	14	7	50,00
2	ErgoGrip	5	4	80,00
3	Spritzenpumpe	14	2	14,29

**Tab. 20: Bei den durchgeführten Studien durch Heuristische Evaluation und Nutzertest gefundene Probleme.**

Weiterhin wurden in dieser Studie die Begleitdokumente evaluiert. Die Beobachtungen während des Tests und die Erfahrungen aus der Praxis zeigen, dass bei Problemen nur selten Begleitdokumente zur Problemlösung verwendet werden. Es ist zwar richtig, dass auch sie bei einer Gebrauchstauglichkeitsprüfung betrachtet werden, jedoch sollte der Hauptfokus immer noch auf der Benutzerschnittstelle liegen.

In dieser Studie wurden wiederum die Aufgabenzeiten als Effizienzmaße gemessen. Da jedoch im Gegensatz zu den vorherigen Studien keine Vergleichsdaten von Wettbewerbsprodukten vorlagen, konnte hierbei auch keine Aussage zur Effizienz in Relation zu Wettbewerbern gemacht werden. Kriterien der Effizienz wurden durch den Testverantwortlichen bestimmt. Hierbei führte er die Aufgaben nach einer Einarbeitungszeit durch und setzte im Falle einer fehlerfreien Aufgabenerledigung die gemessene Zeit als Maßstab. Die Zeiten der Testnutzer, welche Aufgaben ohne Probleme lösen konnten, lagen nahe am Vergleichswert (Abb. 58).



**Abb. 58: Aufgabenbearbeitungsdauer mit Standardabweichung beim Durchschnitt**

Vor dem Nutzertest erhielten die Probanden, welche noch keinen Kontakt mit Spritzenpumpen hatten, durch einen Außendienstmitarbeiter des Herstellers eine Einweisung in den bestimmungsgemäßen Gebrauch. Diese Einweisung macht die Testsituation realistischer, da für die Anwender auch im realen Nutzungskontext entsprechend §5 MPBetreibV eine Einweisung vorgeschrieben ist. Trotz dieser Einweisung zeigten sich bei diesem Test immer noch erhebliche Gebrauchstauglichkeitsprobleme.

Im Gebrauchstauglichkeitstest wurden 14 Probleme gefunden (Abb. 59), welche sich sieben Arbeitsaufgaben der Spritzenpumpe zuordnen lassen. Fraglich war jedoch die Relevanz bzw. der Schweregrad dieser Probleme. Es zeigte sich also auch in dieser Studie die Notwendigkeit eines Bewertungssystems für Gebrauchstauglichkeitsprobleme.

Aufgabe	Problem	EI	Testperson (gelb, wenn Experte)									Σ	Auftrittens- frequenz
			1	2	3	4	5	6	7	8	9		
Spacestation bestücken	Perfusor verkantet, rastet nicht merklich ein											6	66,67
Einschalten	START statt POWER gedrückt											2	22,22
Spritze einlegen	Spritzenwechsel einleiten	X										5	55,56
	Einlegen der Spritze, Verfehlen der Spritzenfixierung											7	77,78
	Leitung eingeklemmt											7	77,78
	Spritze horizontal eingelegt											1	11,11
Rate eingeben und Förderung starten	OK statt START gedrückt	X										5	55,56
	Einzelschritte zur Werteeingabe											5	55,56
Bolus	START statt BOL gedrückt um Bolus zu verabreichen											2	22,22
	OK statt BOL gedrückt											5	55,56
	Im Auswahlmü Bolus gedrückt											1	11,11
	Pfeil zur Rateneinstellung wird zu spät erkannt											3	33,33
Kontrasteinstellung	Im Hauptmenü wird "Scroll Down" nicht erkannt											4	44,44
Global, gesamtes Gerät	Fehlendes Feedback											3	33,33

**Abb. 59: Gebrauchstauglichkeitsprobleme aus Experteninspektion (EI) und Nutzertest mit den dazugehörigen Arbeitsaufgaben. Expertennutzer sind in Gelb hervorgehoben.**

Die Betrachtung der Stichprobengröße und die Berechnung der Fehlerfindungsrate, entsprechend der in Studie 1 beschriebenen Methode, zeigt, dass mit der verwendeten Stichprobengröße von neun Teilnehmern eine Problemfindungsrate von über 97% erreicht werden konnte (Abb. 60).

Innerhalb der Stichprobe dieser Studie konnte zwischen erfahrenen Nutzern (Anästhesiepfleger/Krankenpfleger) und Auszubildenden unterschieden werden. Es war zu erkennen, dass bei den drei erfahrenen Nutzern 12 der 14 Probleme auftraten. Somit war die Problemauftretensrate von 85,7 % bei den Expertennutzern höher als die durchschnittliche Problemfindungsrate von 71,6% bei drei zufälligen Testteilnehmern. Zumindest für diese Studie zeigte sich, dass erfahrene Nutzer nicht unbedingt eine neue Benutzerschnittstelle besser bedienen können. Für die Standardprozedur bedeutet diese Erkenntnis, dass die zu verwendende

Stichprobe aus Repräsentanten der Nutzergruppe mit verschiedenen Erfahrungsgraden bestehen muss.

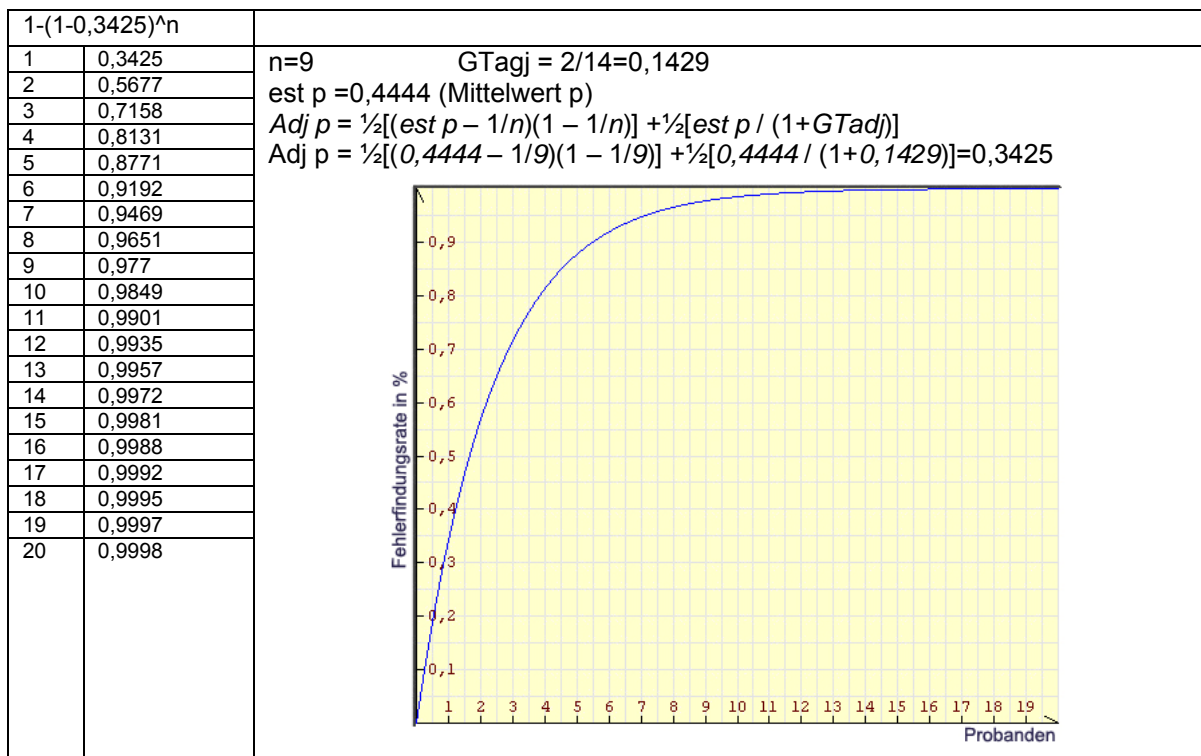


Abb. 60: Berechnung der Fehlerfindungsrate in Abhängigkeit von der Stichprobengröße

Insgesamt zeigte sich, dass die Standardprozedur durch Dritte verstanden wurde und erfolgreich angewendet werden konnte. Der Leitfaden ermöglichte dem Masterstudenten, der wenig Erfahrung im Bereich der Gebrauchstauglichkeitsevaluation hatte, auf strukturierte Weise Gebrauchstauglichkeitsprobleme in der Medizin zu evaluieren.

### **Fazit:**

Es konnten folgende neue Erkenntnisse und Erfahrungen in die Entwicklung der Standardprozedur einfließen:

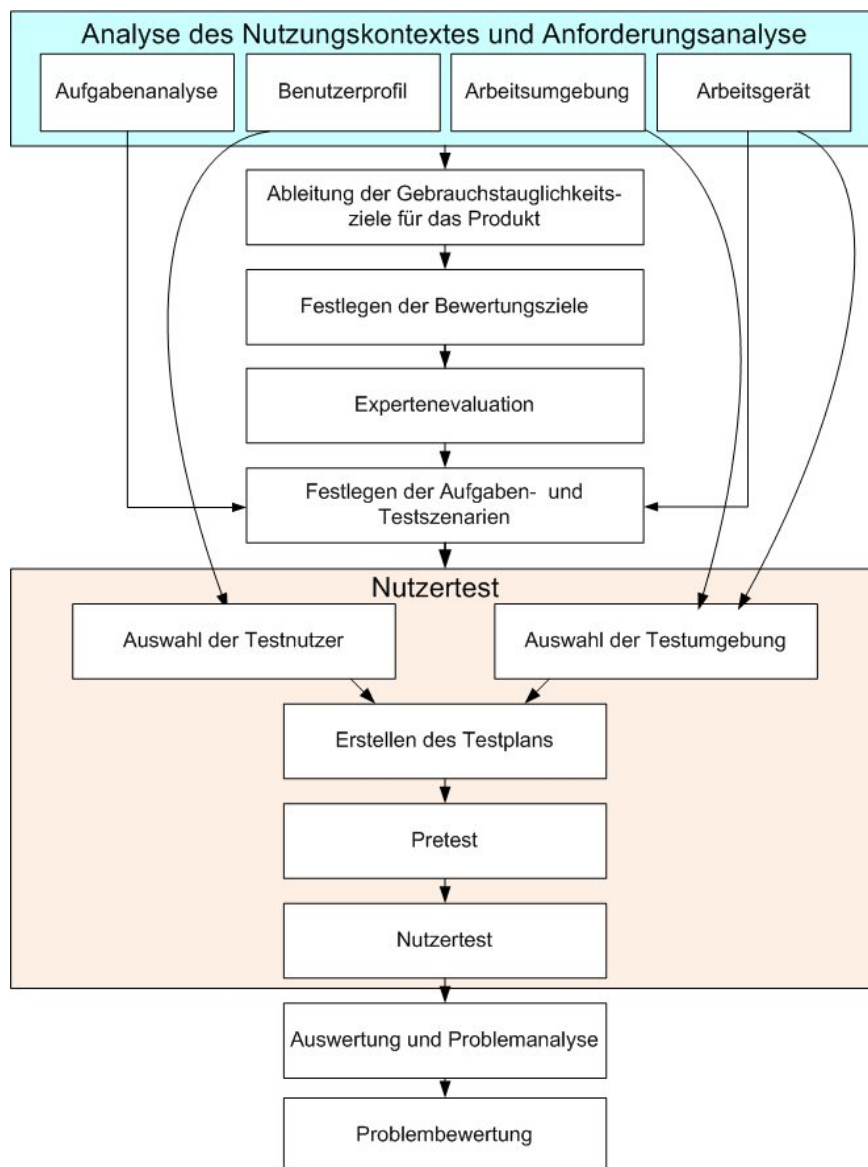
- Für eine Experteninspektion liefern Heuristiken mehr Informationen als die nicht standardisierte Inspektion.
- Die Evaluation der Begleitdokumentation ist nicht der wichtigste Schritt, da diese von den Anwendern meistens nicht gelesen wird.
- Um eine realistische Betrachtung der Benutzerschnittstelle medizinischer Geräte zu gewährleisten, ist die Einweisung entsprechend §5 der MPBetreibV (vgl. Kap. 2.3) notwendig.

- Auch ohne Vergleichsgeräte ist die Gebrauchstauglichkeitsuntersuchung möglich.
- Der Ermittlung eindeutiger Maße bezüglich der Aufgabeneffektivität soll besondere Aufmerksamkeit gewidmet werden. Als Hilfe können Aufgabendurchführung durch den Testverantwortlichen, die Messungen durch Experten, Kriterien bereits durchgeführter Tests, Vergleiche zu bestehenden Systemen oder Performancewerte mit einem Prototyp dienen.
- In dieser Studie erwies sich eine Stichprobengröße von neun Teilnehmern als ausreichend, um eine Problemfindungsrate von über 97,7% (0,91369) zu erreichen. Bereits mit sechs Probanden wäre eine Problemfindungsrate von 91,9 % erreicht worden.
- Erfahrene Nutzer machen nicht unbedingt weniger Fehler als Laien.
- Der Einsatz eines Bewertungssystems für Gebrauchstauglichkeitsprobleme ist notwendig.
- Die Standardprozedur kann durch Dritte durchgeführt werden.

### 4.3 Ergebnisse

Die Ergebnisse der durchgeführten Studien wurden fortwährend in eine Standardprozedur überführt, welche im Folgenden dargestellt ist. Bei den einzelnen Schritten der Standardprozedur werden Beispiele und Anregungen gegeben, welche im praktischen Einsatz als Hilfestellung herangezogen werden können. Je nach medizinischem Gerät können sie angepasst werden.

In der Evaluationsphase wurde klar, dass die in den Studien gewonnenen Ergebnisse nicht ausreichen, um die Gebrauchstauglichkeit medizinischer Geräte zu bewerten. Eine hierzu notwendige Bewertungsmethode wird in den Kapiteln 5 und 6 entwickelt und evaluiert. In einer finalen Studie (► Kap 6.4) wird daraufhin die gesamte Prozedur evaluiert.



**Abb. 61:** Schematische Darstellung der in diesem Kapitel hergeleiteten Standardprozedur

## Standardprozedur zur Bewertung der Gebrauchstauglichkeit medizinischer Geräte

### I Vorbereitungsphase

#### I.a Einarbeitung in das zu untersuchende Objekt

- Verständnis der Art und Weise der Anwendung, Kenntnis über das Produkt (vgl. Kap. 1.2.2.3 u. 4.1.1.2)
- Der Testverantwortliche sollte eine nach §5 Satz 1 MPBetreibV vorgeschriebene Einweisung durch den Betreiber oder den Geräteverantwortlichen erhalten. Der nach §5 Satz 2 MPBetreibV geforderte Ausbildungs- und Qualifikationsgrad ist nicht notwendig (vgl. Kap. 2.3).
- Kenntnis der für das medizinische Gerät geltenden Normen und Bestimmungen (Gesetze, Richtlinien) (vgl. Kap. 2.3)
- Review und Identifizierung aller allgemeinen Designprinzipien für das Produkt (Konsultation von Experten, um relevante Designprinzipien zu erkennen)

#### I.b Anforderungsanalyse

Die Anforderungsanalyse dient dem Verstehen und Festlegen des Nutzungskontextes (vgl. Kap. 4.1.1). Sie besteht aus einer Nutzer-, Aufgaben-, Umgebungs- und Geräteanalyse. Zur Datenerhebung können z.B. Daten der Produkthersteller, Benutzerinterviews, Kontextsitzungen und Nutzerbeobachtungen herangezogen werden.

##### I.b.1 Benutzeranalyse

- Analyse der Nutzer und Erstellung von Nutzerprofilen
- Nutzercharakteristika (Alter, Geschlecht, Kenntnis, Ausbildung etc.)
- Unterscheidung in direkte und indirekte Nutzer

##### I.b.2 Aufgabenanalyse

- Analyse der Nutzerziele, Arbeitsaufgaben und Nutzeraktionen
- Unterscheidung zwischen Primär- und Sekundäraufgaben, sowie Häufigkeit der Nutzung
- Schematische Darstellung (z.B. HTA) (vgl. Kap. 4.1.1.2)
- Welche Aufgaben erfüllen die Nutzer mit dem Produkt, welche Aufgaben können mit dem Produkt generell erfüllt werden?
- Erfassung von Kontext- und Anwendungsszenarien (vgl. Kap. 4.1.5)

##### I.b.3 Analyse der Arbeitsumgebung und der Arbeitsmittel

- Analyse der Arbeitsumgebung (physisch, organisatorisch, sozial, kulturell)
- Analyse der beeinflussenden Arbeitsmittel
- Analyse der leistungsbeeinflussenden Faktoren
- Analyse der Integrierbarkeit, Systemfähigkeit und Kompatibilität (optional)

### **I.c Erstellung von Benutzeranforderungen, funktionalen- und organisatorischen Anforderungen**

Benutzeranforderungen und organisatorische Anforderungen werden neben den funktionalen Anforderungen aus der Betrachtung des Nutzungskontextes (vgl. Kap 4.1.1) erstellt. In einem späteren Schritt werden Ziele mit eventuellen Kompromissen zwischen den Anforderungen abgeleitet. Die Ziele müssen so formuliert werden, dass sie in der folgenden Gebrauchstauglichkeitsuntersuchung evaluiert werden können. Die DIN EN ISO 13407 gibt beispielhaft einige Anforderungen für den Gestaltungsprozess vor:

- Funktions- und Leistungsmerkmale
- relevante gesetzliche Anforderungen
- Sicherheit
- Zusammenarbeit und Informationsaustausch zwischen Benutzern und sonstigen Beteiligten (z.B. Patienten)
- Arbeitsaufgaben und Ziele der betroffenen Nutzer
- Ausführung der Arbeitsaufgabe
- Arbeitsgestaltung und Organisation
- Durchführbarkeit des Betriebes und der Wartung
- Hygiene und Wiederaufbereitung
- Gestaltung der Mensch-Maschine-Schnittstelle und des Arbeitsplatzes

Ein Leitfaden zur Erstellung von Nutzungsanforderungen ist im Prüfhandbuch zur Gebrauchstauglichkeit der Deutschen Akkreditierungsstelle Technik dargestellt [77].

### **I.d Ableiten und Operationalisieren der Gebrauchstauglichkeitsziele für das Produkt**

Aus den evaluierten, funktionalen und organisatorischen Anforderungen sowie den Anforderungen der Nutzer an das Produkt müssen unter Zuhilfenahme der Ergebnisse aus der Benutzer-, Aufgaben- und Umfeldanalyse qualitative und quantitative Gebrauchstauglichkeitsziele extrahiert werden, welche das Produkt in seinem Nutzungskontext erfüllen sollte. Sie sind in enger Zusammenarbeit mit den Nutzern und Herstellern zu entwickeln.

Diese Ziele sollten die Anforderungen der DIN EN ISO 13407 erfüllen und somit:

- eine deutliche Darstellung der benutzerorientierten Gestaltungsziele liefern,
- auf den relevanten Nutzungskontext bezogen sein,
- geeignete Prioritäten für die verschiedenen Anforderungen setzen,
- messbar sein, d.h. messbare Kriterien liefern, anhand derer die entstehende Gestaltungslösung evaluiert werden kann,
- durch Benutzer oder deren Interessensvertreter bestätigungsfähig sein,
- vorgeschriebene gesetzliche Anforderungen einschließen und
- dokumentiert sein.



Die Ziele, welche das Produkt im Hinblick auf die Benutzer- und die funktionalen- bzw. organisatorischen Anforderungen erfüllen soll, dienen als Leitfaden, um die Gebrauchstauglichkeit des Produktes zu evaluieren und zu bewerten. Diese Ziele müssen operationalisiert, d.h. in messbare Ziele überführt werden (vgl. Kap. 2.4). Sie werden in Form von expliziten Konzeptionsspezifikationen, die subjektive oder objektive Leistungsmaße beinhalten können, ausgedrückt.

Gebrauchstauglichkeitsziele sollen klar und unmissverständlich formuliert sein. Sie sollen Leistungen messen und die Möglichkeit bieten, Vergleiche zu anderen Produkten oder Entwürfen herzustellen. Weiterhin sollen sie durch einen Nutzertest überprüfbar sein.

Grundlegende Gebrauchstauglichkeitsziele lassen sich der DIN EN 9241-11 oder der DIN EN IEC 60601-1-6 entnehmen. Als weitere Hilfe für die Zielerstellung können die von Hölscher et al. entwickelten Prüfpunkte zur Gebrauchstauglichkeit verwendet werden, welche ein breites Spektrum relevanter Prüfvorgaben für medizinische Geräte aufzeigen (vgl. Kap. 3.2.1).

Beispiele:

	Ergebnis aus der Anforderungsanalyse	Gebrauchstauglichkeitsziel	Messbare Gebrauchstauglichkeitsziele
Beispiel 1	Die Nutzer wollen ein Rezept mit dem neuen mobilen Gerät genauso schnell und einfach erstellen wie vorher per Hand.	Die mobilelektronische Eingabe eines Rezeptes soll genauso schnell wie die Eingabe per Hand erfolgen. Die Nutzer sollen mit der Eingabe zufrieden sein.	<ul style="list-style-type: none"> <li>- 100 % der Nutzer müssen ein Rezept mit dem mobilen Gerät nach einer 10-minütigen Trainingssession eingeben können.</li> <li>- Die Eingabezeit eines vorgegeben Textes soll <math>\leq</math> der Eingabe per Hand sein.</li> <li>- 8 von 10 Nutzern sollen auf einer Skala von 1 (schlecht) bis 5 (exzellent) die Eingabe mit 4 oder 5 bewerten.</li> </ul>
Beispiel 2	Die Nutzungskontextanalyse ergab, dass drei Personen im OP einen Bildschirm beobachten müssen.	Die Anzeige muss sowohl im Nahbereich (1m Entfernung), als auch aus 3 Metern Entfernung gut sichtbar sein (keine Verdeckung durch eine Person). Alle Personen müssen in der Lage sein, Farben zu erkennen und den Text zu lesen.	<ul style="list-style-type: none"> <li>- 100% der Nutzer müssen jederzeit während des laufenden OP-Betriebes Texte lesen und Farben erkennen können.</li> </ul>
Beispiel 3	Die Anwender finden, dass sie mit dem derzeitigen Gerät aufgrund der schlechten Menüführung für Arbeitsaufgaben zuviel Zeit benötigen (derzeitige Bearbeitungszeit: $\varnothing$ 4:20 min).	Die Aufgabenerfüllung muss wesentlich schneller sein als bisher, und die Anwender müssen mit der Aufgabenerfüllung zufrieden sein.	<ul style="list-style-type: none"> <li>- 80% der Anwender müssen die Arbeitsaufgaben in weniger als 3:30 min. erledigen können.</li> <li>- Der Grad der durchschnittlichen subjektiven Zufriedenheit mit der neuen Menüführung muss um einen Punkt besser sein als bisher.</li> </ul>

### I.e Festlegung von Bewertungszielen / Messungen

Die Messung der Gebrauchstauglichkeit beruht auf der Ermittlung der Gebrauchstauglichkeitsmaße Effektivität, Effizienz und Zufriedenheit (vgl. Kap. 2.4), welche aus Interaktionsdaten der Benutzer mit dem Gerät extrahiert werden. Die in „I.d.“ ermittelten Gebrauchstauglichkeitsziele werden hierbei den Maßen der Gebrauchstauglichkeit zugewiesen. Hierbei können objektive Daten (z.B. durch Messung der Arbeitsgeschwindigkeit) und subjektive Daten (Benutzermeinung) erhoben werden. Objektive Daten liefern direkte Ergebnisse bzgl. Effektivität und Effizienz, subjektive Daten stehen im direkten Zusammenhang mit Zufriedenheit. Die im medizinischen Kontext elementare Sicherheit wird durch die Maße der Gebrauchstauglichkeit, gewichtet mit einem Risikoindex, ermittelt.

	Gebrauchstauglichkeitsziel	Maße der Effektivität	Maße der Effizienz	Maße der Zufriedenheit
Beispiel 1	Die mobilelektronische Eingabe eines Rezeptes soll genau so schnell wie eine Eingabe per Hand erfolgen, und die Nutzer sollen mit der Eingabe zufrieden sein.	100 % der Nutzer müssen ein Rezept mit dem mobilen Gerät nach einer 10-minütigen Trainingssession eingeben können.	Die Eingabezeit eines vorgegeben Textes soll $\leq$ der Eingabe per Hand sein.	8 von 10 Nutzern sollen auf einer Skala von 1 (schlecht) bis 5 (exzellent) die Eingabe mit 4 oder 5 bewerten.
Beispiel 2	Die Anzeige muss sowohl im Nahbereich (1m Entfernung), als auch aus 3 Metern Entfernung gut sichtbar sein (keine Verdeckung durch eine Person). Alle Personen müssen in der Lage sein, Farben zu erkennen und den Text zu lesen.	100% der Nutzer müssen jederzeit während des laufenden OP-Betriebes Texte in Entfernungen von 1m bis 3m lesen und Farben erkennen können.	Die Lesegeschwindigkeit in einer Entfernung von 3 m darf nicht viel niedriger sein, als in einer Entfernung von 1m	Alle Testnutzer sollen auf einer Skala von 1 (schlecht) bis 5 (exzellent) den Lesekomfort mit 4 oder 5 bewerten.

### I.f Heuristische Evaluation oder Expertenreview

Um erste Erkenntnisse bezüglich der Gebrauchstauglichkeit zu evaluieren, ist eine heuristische Evaluation zu bevorzugen (vgl. Kap. 2.4.1.1, Kap. 4.2), wenigstens eine Expertenevaluation sollte durchgeführt werden.

Wenn möglich, sollte auch ein Review über bestehende Vorkommismeldungen mit dem Gerät durchgeführt werden, um Ursachen für mögliche Probleme, Zwischenfälle bzw. Beinahezweischenfälle aufzudecken und zu erhärten. Voraussetzung hierfür ist der Zugang zu einer entsprechenden Datenbank. Viele Hersteller, Krankenhäuser oder auch das BfArM verfügen über solche Datenbanken.

In dieser Phase sollte auch der vorhersehbare Missbrauch berücksichtigt werden.

### **I.g Festlegen der Aufgaben- und Testszenarien**

Die im Nutzertest gestellten Aufgaben sollten den Arbeitsaufgaben der realen Nutzungsumgebung entsprechen (vgl. Kap. 4.1.5). In den meisten Fällen wird die komplette Abarbeitung aller Nutzeraufgaben nicht möglich sein. In diesem Fall sollte die Auswahl der Aufgaben mit besonderer Sorgfalt geschehen, da sie erheblichen Einfluss auf die Testresultate haben kann. Auch im Sinne der Reliabilität (vgl. Kap. 4.1.8) sind bei der Selektion der Aufgaben folgende Punkte zu beachten:

- möglichst alle Primäraufgaben und die wichtigsten Sekundäraufgaben
- Aufgaben, welche sich in der Expertenevaluation als problematisch herausgestellt haben
- Sollten bereits sicherheitskritische Situationen oder Arbeitsschritte in der Anforderungsanalyse erkannt werden, so sind Aufgaben zu wählen, die diese überprüfen.
- Aufgaben, welche vom Gerätehersteller, Betreiber und/oder Kunden/Anwender vorgeschlagen werden

Die gestellten Aufgaben sollten in Szenarien eingebettet sein (vgl. Kap. 4.1.5), um dem Nutzer einen realistischeren Zugang zu ihnen zu verschaffen. In einem Szenario gibt man dem Nutzer ein Ziel und weitere Informationen zur Zielerreichung vor, jedoch keine Schritte zur Aufgabenlösung. Typischerweise wird geprüft, ob der Nutzer selbständig sein Ziel erreichen kann. Ein gutes Szenario ist kurz, in der Sprache der Nutzer formuliert und eindeutig, sodass es jeder Nutzer versteht. Es soll dem Testnutzer genügend Informationen zur Aufgabenerledigung geben und direkt mit den Aufgaben und der Testintention verbunden sein (vgl. Kap. 4.1.5).

In welcher Form die in der DIN EN 9241-110 beschriebenen Gestaltungsgrundsätze (vgl. Kap. 2.4) zu berücksichtigen sind, hängt von den jeweiligen Anforderungen der Nutzer, der Nutzungsumgebung und der Arbeitsaufgaben ab. Hierbei sollen besonders die Merkmale der Aufgaben berücksichtigt werden, die bestimmte Klassen der Benutzergruppe ausführen müssen (z.B. Individualisierbarkeit für Expertennutzer und Lernförderlichkeit für Anfänger).

Weiterhin sollten für die Aufgaben eindeutige Maße bezüglich der Aufgabeneffektivität definiert werden (vgl. Kap 4.1.2). Diesem Schritt soll, soweit keine Vergleichsdaten von Wettbewerbern vorliegen, besondere Aufmerksamkeit gewidmet werden. Als Hilfe dienen folgende Richtlinien:

- Der Testverantwortliche führt die Aufgabe durch und setzt die Kriterien,
- Messungen von Experten gelten als Maßstab,
- Kriterien von bereits durchgeführten Tests,
- vereinbarte Werte zwischen den Verantwortlichen,
- Vergleiche zu bestehenden Systemen oder Performancewerte mit einem Prototyp.

## II. Gebrauchstauglichkeits-Test

### II.a Auswahl und Akquise der Testnutzer

- Repräsentative Stichprobe entsprechend dem Benutzerprofil (vgl. Kap. 4.1.4)
- Entscheidung, ob zusätzlich zu den direkten Nutzern auch indirekte Nutzer einbezogen werden
- Berücksichtigung verschiedener Nutzergruppen und Erfahrungen
- Zu einer Bewertung sind mindestens 9 Teilnehmer erforderlich, bei einer heterogenen Nutzergruppe sollte die Population entsprechend höher sein (vgl. Kap. 4.1.4).
- Für medizinische Geräte können zur Repräsentation neuer und unerfahrener Nutzer sowohl Auszubildende als auch Studenten höherer Semester der entsprechenden Fachrichtung herangezogen werden, da diese als zukünftige Nutzer anzusehen sind.

In der Praxis ist es relativ schwierig, geeignete Nutzer für Gebrauchstauglichkeitstests zu identifizieren und für einen Test zu motivieren. Daher sollte frühzeitig ein Probandennetzwerk aufgebaut und mit der Rekrutierung begonnen werden. In den meisten Fällen kann der Hersteller des medizinischen Gerätes bei der Rekrutierung unterstützende Hilfe leisten. Hierbei sollte jedoch auf potentielle Lobbyisten geachtet werden.

### II.b Auswahl der Testumgebung

Aufbau einer realitätsgetreuen Testumgebung entsprechend der Umfeldanalyse; Aufbau der entsprechenden Beobachtungs-, Mess- und Aufzeichnungstechnik gemäß dem Interessensfokus; Aufbau entsprechender zusätzlich relevanter Hardware (Eye-tracking, EMG etc.) (vgl. Kap. 4.1.6); eventuell sollten entsprechende Einrichtungen für Stresstests vorbereitet werden.

### II.c Erstellen der Testunterlage

Die Testunterlage ist ein Dokument oder eine Sammlung von Dokumenten, die den reibungslosen und immer konstanten Ablauf einer Testreihe garantieren. Elemente der Testunterlage sind:

- Evtl. eine kurze Einführung in das Produkt (kann aber auch mündlich erfolgen)
- Verschwiegenheitserklärung
- Einverständniserklärung für Audio- und Videoaufzeichnung
- Festhalten der Aufgabenszenarien, welche die Testperson während der Durchführung des Tests begleiten
- Dokument, in dem die ggf. anzubietende Hilfestellung für jede Testaufgabe im Detail festgelegt ist
- Fragebögen während des Tests (ASQ)
- Fragebögen nach dem Test (z.B. Vergleichsfragebögen, Fragen zu einzelnen Aktionen, PSSUQ und ISOMetrics)
- Demografischer Fragebogen
- Interviewleitfaden
- Vergleichsfragebögen
- Beobachtungsprotokoll

## II.d Erstellung des Testplans

Für die Gebrauchstauglichkeitsuntersuchung wird ein Prüfprogramm erstellt, aus dem hervorgeht, welche Eigenschaften untersucht werden und welche Methoden zum Einsatz kommen. Dieses Programm (Testkonzept) muss unmissverständlich geschrieben werden, und es muss klar daraus hervorgehen, was und wie gemessen wird. Mindestinhalte sollten sein [105]:

- Kontaktdaten des Testverantwortlichen bzw. der Institution
- Beschreibung des zu beurteilenden Systems
- Ergebnisse der Anforderungsanalyse
- die benutzerorientierten Prüfziele mit den entsprechenden Gebrauchstauglichkeitsmaßen
- Beschreibung der zu testenden Aufgaben mit dem entsprechenden Testziel und Gebrauchstauglichkeitsmaßen sowie eventuellen Abbruchkriterien und Hilfestellungen
- Methoden der Evaluation
- Beschreibung der Stichprobe
- Beschreibung der Testumgebung
- Beschreibung des Testablaufs
- Terminliche Festlegung

Die Inhalte des Testplans können auch für die Dokumentation verwendet werden.

Szenario	Aufgaben	Testziel	Testmaße			Hilfelevel	Sinn der Aufgabe?
			Effizienz	Effektivität	Zufriedenheit		
Szenario 1: Behandlung beginnen	Nehmen Sie das Gerät in Betrieb	Kann das Gerät in Betrieb genom- men werden?	Grad der Zielerreichung (in Prozent)  Prozentsatz der Benutzer, die Aufgaben erfolgreich abschließen  Durchschnittliche Genauigkeit der abgeschlossenen Aufgaben  Fehlerhäufigkeit  Häufigkeit der Unterstützung	Zeit für die Erledigung einer Aufgabe  Abgeschlossene Aufgaben pro Zeiteinheit	Einstufungsskalen für Zufriedenheit  Häufigkeit von Beschwerden  Subjektive Beobachtung  Fragebogenfragen mit spezieller Relevanz	Keine Hilfe, jedoch Handbuch und Telefon stehen bereit	Aufgabe ist eine Primäraufgabe, und in der Experten-evaluation wurde vermutet, dass der Einschalter nur schwer gefunden wird.

## II.e Vortest

Kein Nutzertest sollte ohne einen Vortest stattfinden. In diesem wird das Testdesign überprüft. Besonderes Augenmerk wird auf die Aufgabenszenarien, den natürlichen Ablauf, die Aufnahmetechnik, die Vollständigkeit des Instrumentariums und die Kamerakalibrierung gelegt. Nach dem Vortest wird das Testdesign angepasst und ggf. ein neuer Vortest durchgeführt.

## II.f Gebrauchstauglichkeits-Test

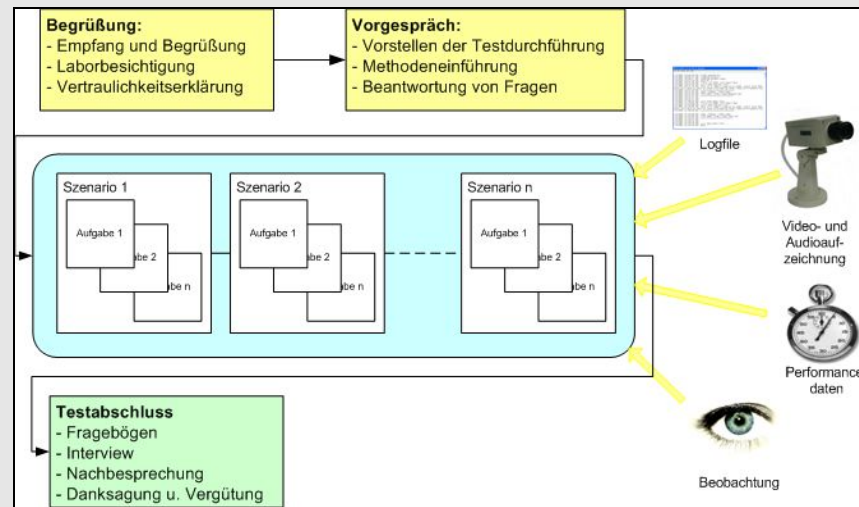


Abb. I: Ablauf eines Gebrauchstauglichkeitstests

### 1. Einführung

- Begrüßung der Testteilnehmer und kurze Laborbesichtigung
- Unterzeichnung der Vertraulichkeitserklärung und der Bereitschaft zur Videoaufzeichnung
- Erklärung der Testprozedur, der Methode: „Thinking Aloud“ und ggf. Einweisung in die Geräte
- Erinnerung: „Die Geräte werden getestet, nicht Sie.“
- Klärung offener Fragen

### 2. Testdurchführung

- Durchführung der Aufgabenszenarien
- Protokollierung von Auffälligkeiten und Videoaufzeichnung
- Fragebogen nach der Bearbeitung der einzelnen Aufgaben (ASQ)
- Mitstoppen der Bearbeitungszeit je Szenario, je Aufgabe und Teilaufgabe
- Der Testleiter sollte primär eine passive Rolle einnehmen.

Während des Tests sollte auf eine entspannte Atmosphäre zwischen dem Probanden und dem Evaluator geachtet werden. Dennoch sollte auch die typische Nutzersituation simuliert werden. Wird das zu testende Gerät beispielsweise in Stresssituationen verwendet, so sollte auch Stress provoziert werden. Auch können Störungen simuliert werden, wenn dadurch bestimmte Funktionen des Gerätes geprüft werden können.

#### Regeln für das Einschreiten:

Einschreiten, wenn...

- ...das Gerät fehlerhaft ist,
- ...sich ein ungeplantes Szenario anbahnt, welches nicht in das Testkonzept passt,
- ...die Testperson hoffnungslos überfordert ist - in diesem Fall sollten Sie der Testperson behilflich sein, an den Punkt zurückzukommen, von dem an sie sich in der Gerätebedienung „verlaufen“ hat.

Das Einschreiten sollte nicht in Form von direkten, sondern von erklärenden Fragen erfolgen:

- „Können Sie mir bitte die Schritte erklären, die Sie soeben gemacht haben?“
- „Welche Aktion haben Sie bei der Auswahl dieses Menüeintrags erwartet?“

### 3. Testabschluss

- Abschließendes strukturiertes Interview und/oder Fragebogen. Als geeignete Fragebogenelemente erwiesen sich ASQ und PSSUQ (vgl. Kap. 2.4.1.2), bei deduktivem Ansatz eignen sich Vergleichsfragebögen
- Demografischer Fragebogen
- Ggf. Vorführen von Highlightvideos und Interview
- Beantwortung offener Fragen
- Danksagung

## III. Auswertung / Dokumentation

### III.a Datenanalyse

- Transkription des Befragungsmaterials
- Auswertung der subjektiven und objektiven Daten (z.B. Aufgabenzeit, Fehlerraten, Nutzerreaktion, Fragebögen, Nutzeraussagen)
- Auswertung der Daten für Effektivität, Effizienz und Zufriedenheit
- Statistische Auswertung

### III.b Gebrauchstauglichkeitsproblemanalyse

- Fehlerklassifizierung (vgl. Kap. 5.1, 5.2)
- Fehler- und Problemanalyse (vgl. Kap. 5.4)
- Rückschlüsse auf Gebrauchstauglichkeitsprobleme

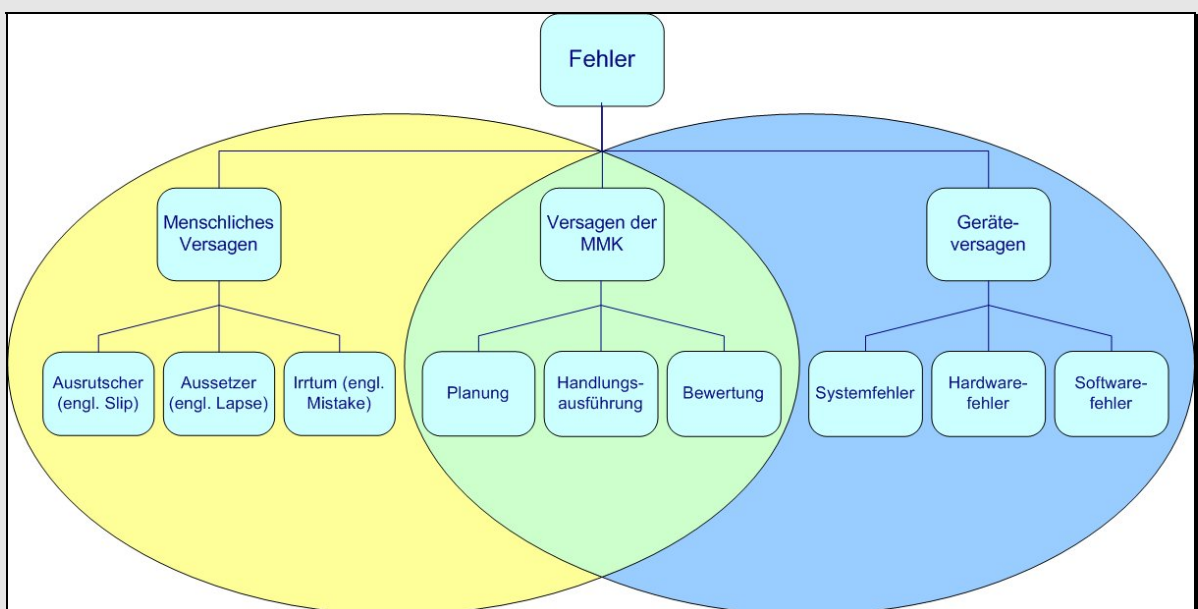


Abb.II: Kategorisierung von Gebrauchstauglichkeitsproblemen

Maschinenfehler (Hardwarefehler) sind keine Fehler, welche auf Gebrauchstauglichkeitsprobleme zurückzuführen sind, und sollten normalerweise bereits durch die Qualitätsprüfung des Herstellers abgedeckt werden. Bei Auftreten in der Gebrauchstauglichkeitsevaluation sollten sie jedoch doku-

mentiert werden. Rein menschliche Fehler werden nur insofern betrachtet, als dass die Geräte Sicherheitsroutinen aufweisen sollten, welche menschliche Fehler vermeiden und die Auswirkungen mindern etc..

### **III.c Problembewertung**

- Bewertung der Gebrauchstauglichkeitsprobleme mit einem Bewertungsschema (vgl. Kap. 6.3), durch Problemeinstufung, Auftretenswahrscheinlichkeit und Auswirkungen
- Abwägung zwischen Risiko und Nutzen (vgl. Kap. 5.4.3)
- Prüfung, ob das Problem mit dem derzeitigen Stand von Technik und Wissenschaft behebbar ist

### **III.d Herausarbeiten von Ansätzen für die Produktverbesserung**

Auch bei einer reinen Bewertung ist dieser Schritt auszuführen. Er garantiert, dass der derzeitige Stand von Technik und Wissenschaft geprüft wird, um die Probleme zu beheben. Sind Probleme mit derzeitigem Wissen und aktueller Technik nicht zu beheben, so ist eine Risiko-Nutzenabwägung durchzuführen.

### **III.e Abschlussbericht**

Ein standardisierter Abschlussbericht (vgl. Kap. 4.1.7) sichert Reliabilität und Vollständigkeit. Er sollte auch entsprechende Highlightvideos der Tests beinhalten, welche die Ergebnisse belegen.



## **5 Grundlagen und Rahmenbedingungen für die Entwicklung einer Bewertungsmethode der Gebrauchstauglichkeit medizinischer Produkte**

In Kapitel 4 wurde eine Evaluationsmethode entwickelt, mit der möglichst viele Gebrauchstauglichkeitsmerkmale medizinischer Geräte evaluiert werden können. Es wurde festgestellt, dass es nicht ausreicht, diese lediglich zu identifizieren. Für eine Beurteilung der Gebrauchstauglichkeit müssen die Merkmale konkretisiert werden.

Bisher gibt es noch kein Verfahren, um Gebrauchstauglichkeitsprobleme in der Medizin zu bewerten. Die bisherigen Bewertungssysteme hierfür sind allgemein ausgerichtet und berücksichtigen nicht ausreichend die sicherheitskritischen Aspekte des medizinischen Arbeitssystems. Gebrauchstauglichkeitsprobleme sind jedoch häufig Ursachen für Fehler in diesem System (vgl. Kap. 2.2.2.2). Dennoch können die bisher bekannten Fehlerklassifikationsmodelle oder derzeitigen arbeitswissenschaftlichen Analyseverfahren nicht angewendet werden, da sie in zu geringem Masse die interaktive Komponente der Mensch-Maschine-Kommunikation berücksichtigen. Dies ist ebenfalls bei der Risikobetrachtung der Fall. Lediglich die bereits beschriebene PROMEDIKS Methode (vgl. Kap. 3.2) von Backhaus beschäftigt sich primär mit der Betrachtung von Gebrauchstauglichkeit im medizinischen Kontext. Allerdings eignet sich diese Methode eher für eine entwicklungsbegleitende Gebrauchstauglichkeitsbetrachtung und ist daher weniger auf einen bewertenden Charakter ausgerichtet.

Dieses Kapitel beschäftigt sich mit der Entwicklung und Evaluation einer Methode zur Bewertung der Gebrauchstauglichkeitsmerkmale. Ziel hierbei ist es einerseits, Entwicklern und Herstellern die Priorität der gefundenen Probleme aufzuzeigen und andererseits, Anwendern und Betreibern medizinischer Geräte Informationen zu deren Gebrauchstauglichkeit zur Verfügung zu stellen. Die Ergebnisse dieser Methode sollen zur Vergabe eines Gebrauchstauglichkeitsnachweises dienen.

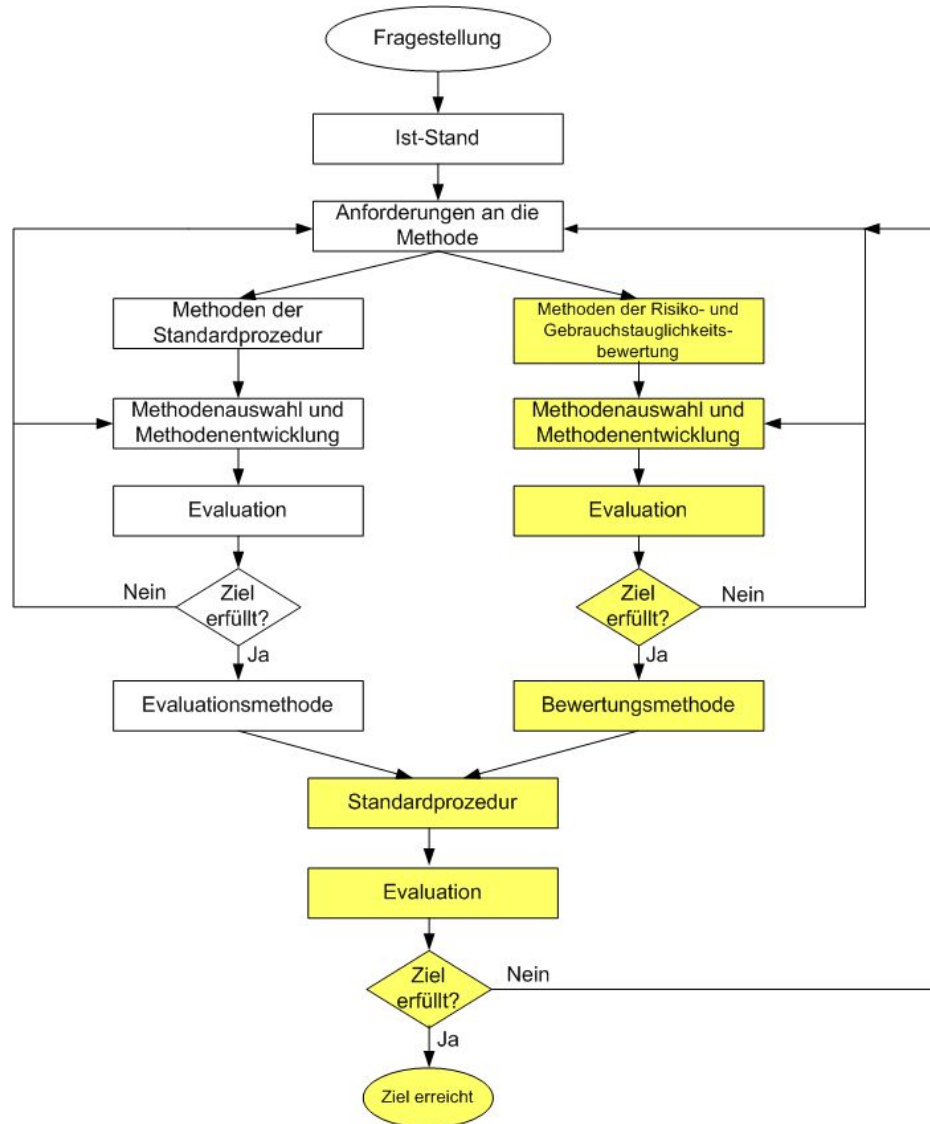
Für diese, die Evaluationsmethode ergänzende Methode, sollen auch die in Kapitel 3.3 beschriebenen Anforderungen gelten. Besonders hervorzuheben ist hierbei, dass sie die speziellen Anforderungen des sicherheitskritischen Arbeitssystems bei der Bewertung berücksichtigt. Den von Schmidkte erhobenen Anforderungen für Bewertungskriterien muss ebenfalls Folge geleistet werden. Hierzu zählen: Reliabilität, Validität, Objektivität, Ökonomie und Nützlichkeit [358]. Die

Methode soll die aus Arbeitssystem, Arbeitsplatz oder Arbeitsmitteln hervorgehenden und im Umfeld wirksamen Belastungsfaktoren erfassen und bewerten. Sie soll auf nachvollziehbaren, theoretischen Modellen aufbauen, welche dem Hersteller, Anwender oder Betreiber medizinischer Geräte eine Ergebnisinterpretation möglich machen.

Weiterhin ist bei der Methode der pragmatische Charakter hervorzuheben. Innerhalb eines angemessenen Kosten-Nutzen-Rahmens wird bewusst auf eine vollständige Erhebung aller Merkmale verzichtet. In Kapitel 4.2 wurde gezeigt, dass mit neun Testnutzern über 90% der Probleme einer Schnittstelle gefunden werden können. Mit diesen Voraussetzungen soll die Methode eine Bewertung durchführen. Auch sollen durch diese Methode nur die Merkmale ausgewählt werden, welche für den Hauptanwenderkreis des medizinischen Gerätes und der Sicherheit bedeutsam sind. Merkmale, welche z.B. das Serviceteam betreffen, werden somit nicht betrachtet.

Die Methode soll einen Beitrag zur Optimierung des medizinischen Arbeitssystems (vgl. Kap. 1.2.2.1) liefern. Sie soll Anwendern, Betreibern, Einkäufern und Herstellern medizinischer Geräte zeigen, welche medizinischen Produkte gebrauchstauglich sind, und welche Mängel und positiven Eigenschaften sie unter Umständen aufweisen.

Im Folgenden werden zuerst die theoretischen Rahmenbedingungen und Klassifizierungsschemata für Fehler vorgestellt, um anschließend aus den Grundlagen und Methoden des Risikomanagements und des Usability Engineerings in einem iterativen Prozess mit abwechselnden Entwicklungs- und Evaluationsphasen eine Methode zur Bewertung von Gebrauchstauglichkeitsproblemen in der Medizin zu erarbeiten (Abb. 62). Es werden zuerst die verschiedenen Fehlerarten erfasst, um ein Fehlerklassifikationssystem für die Fragestellung dieser Arbeit herzuleiten. Danach wird, ausgehend von den theoretischen Grundlagen der Sicherheit und der Risikobetrachtung sowie der Fehlerbewertung in der Mensch-Maschine-Kommunikation, ein Bewertungsmaßstab für Gebrauchstauglichkeitsmerkmale bei medizinischen Geräten entwickelt.



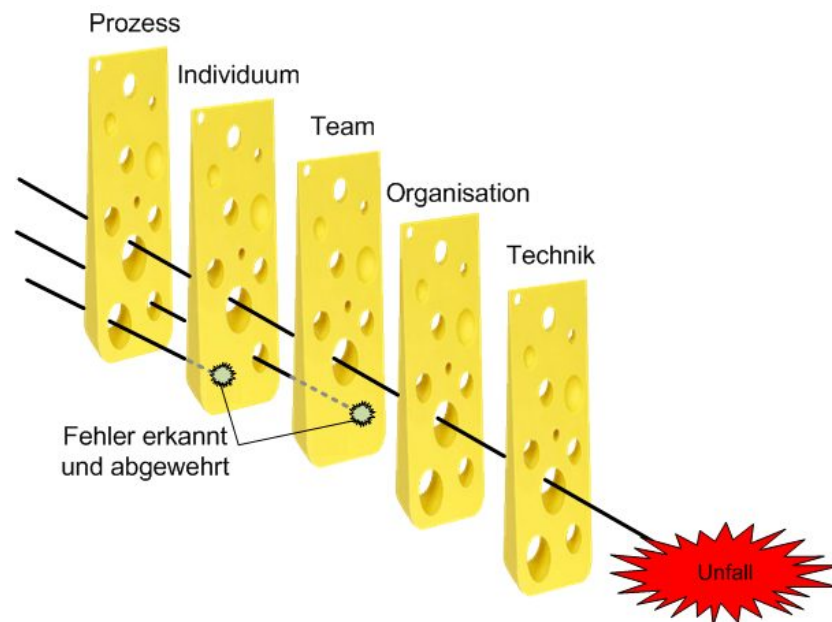
**Abb. 62: Entwicklungsprozess der Standardprozedur. Der in diesem Kapitel entwickelte Bewertungsteil und die finale Evaluation sind markiert.**

## 5.1 Fehler und Fehlerklassifikation

Gebrauchstauglichkeitsprobleme zeichnen sich durch die aus ihnen entstehenden Fehler aus. Um Fehler und deren Ursache zu erfassen, müssen diese zuerst klassifiziert werden. Die Fehlerklassifizierung ist ein Vorgang, um Fehler in Kategorien zu unterteilen. Im folgenden Kapitel werden Klassifizierungssysteme für Maschinenfehler, menschliche Fehler, Fehler in sicherheitskritischen Systemen, Fehler aufgrund von Gebrauchstauglichkeitsproblemen und medizinische Fehler betrachtet und auf ihre Relevanz für die in dieser Arbeit betrachtete Fragestellung überprüft. Eine anschließende Klassifikation für Fehler, hervorgerufen durch

Gebrauchstauglichkeitsprobleme in der Medizintechnik, kann somit aufgestellt werden.

Fehler treten unvermeidlich in allen Systemen auf. Weder vollkommene Perfektion im Design noch absolute Sicherheit gegen Fehler sind erreichbar. Auch muss ein Fehler nicht unbedingt zu einem Zwischenfall führen. Oft bestehen Fehler in einem System, welches den Anwender gar nicht erreicht. Reason beschreibt dieses Phänomen der latenten Fehler mit einem „Schweizer-Käse-Modell“. Die Scheiben des Käses fungieren hierbei als „Schutzschilde“ (Abb. 63). Erst wenn alle „Schutzschilde“ durchdrungen werden, tritt der Fehler auf [332].



**Abb. 63: Schweizer-Käse-Modell nach Reason, bezogen auf das medizinische Arbeitssystem [332]**

Bei der Gebrauchstauglichkeitsbetrachtung liegt der Fokus auf Gebrauchstauglichkeitsproblemen der Maschine (des Gerätes), bzw. auf menschlichen Fehlern, hervorgerufen durch die Maschine oder durch die Kommunikation mit der Maschine. Um jedoch die gesamte Problematik zu erfassen, ist es unabdingbar, sowohl menschliche Fehler, Fehlhandlungen und das Zusammenwirken zwischen menschlichen Fehlern und dem System als auch das Zusammenwirken zwischen menschlichen Fehlern und Gebrauchstauglichkeitsproblemen des Systems zu betrachten.

Um sich der Problematik einer Klassifikation von Fehlern, ausgelöst durch Gebrauchstauglichkeitsprobleme in der Medizin, zu nähern, muss zuerst der Begriff „Fehler“ definiert werden. Maschinenfehler sind auf einfache Weise durch einen Defekt des Systems [380] zu erklären. Mathematisch handelt es sich um die

Abweichung ( $x_i$ ) von einem erwarteten Wert ( $x$ ). Nach der DIN EN 8402 ist ein Fehler gleichzusetzen mit der „Nichterfüllung einer festgelegten Forderung“ [111]. Maschinenfehler lassen sich durch den Einsatz hochwertiger Bauteile verringern. Ein hoher MTBF-Wert<sup>42</sup> (Mean Time Between Failures) verringert hierbei die Auftretenswahrscheinlichkeit von technischen Fehlern. Dennoch tritt nach Leape die Mehrzahl der Fehler im engen Zusammenhang mit Handlungen der Bediener auf [229]. Derartige Fehler sind wesentlich komplexer. Um sie zu erfassen, ist eine komplette und genaue Analyse der Mensch-Maschine-Interaktion notwendig. Im Folgenden werden die für diese Arbeit wichtigen Bereiche beschrieben.

### 5.1.1 Menschliche Fehler

Generell muss man bei menschlichen Fehlern zwischen einem Fehler als Handlung, also dem Begehen eines Fehlers, und dem Fehler als Ergebnis einer Handlung unterscheiden. Beispielhafte Definitionen für letztere Möglichkeit verstehen unter einem Fehler „ein nicht beabsichtigtes, oft auch nicht erwartetes, unerwünschtes Ergebnis einer bewusst oder unbewusst ausgeführten oder unterlassenen Maßnahme“ [322] oder eine „Abweichung der Ergebnisse zielgerichteter menschlicher Handlungen von den Zielen der Handlung“ [32]. Menschliche Fehler als Handlung werden beispielsweise von Hagen & Mays formal definiert als „menschliches Versagen in einer definierten Aufgabe (oder Ausführung einer verbotenen Aufgabe) innerhalb einer spezifizierten Ausführungsgenauigkeit, Reihenfolge oder Zeit, welche in einer Beschädigung von Material und Besitz oder Störung geplanter Abläufe münden kann“ [161]. Die DIN EN 61508-4 beschreibt menschliches Versagen (engl.: human error) und Irrtum (engl.: mistake) als „Handlung oder Unterlassung eines Menschen, die zu einem unerwünschten Ergebnis führt“ [101].

Die genannten Definitionen sehen den Fehler als einen defizitären Teil oder das Ergebnis einer ansonsten richtigen Handlung. Allerdings können Fehler bereits im Vorfeld der Handlung auftreten. Daher wird ein Fehler häufig wie folgt verstanden: Ein richtiges Vorhaben wird nicht wie geplant durchgeführt, oder dem Geschehen liegt ein falscher Plan zugrunde [220, 231, 299, 332, 388]. Diese Definition, welche Grundlage dieser Arbeit ist, umfasst neben dem defizitären Teil der Handlung auch die Phase vor der eigentlichen Handlung, d.h. die Phase, in der ein Fehler

---

<sup>42</sup> MTBF-Wert: zu erwartende mittlere Zeit, welche ein System oder Bauteil ohne Fehler arbeitet

entstehen kann. Sie beschreibt somit den Fehler als Teilmenge eines umfassenden Konzeptes.

Als Ursache menschlicher Fehler sind verschiedene, sowohl extrinsische als auch intrinsische Faktoren zu nennen. Intrinsische, fehlerbeeinflussende Faktoren sind unabhängig von äußeren Umständen oder der Arbeitsaufgabe, wohingegen extrinsische Faktoren durch die Umwelt beeinflussbar sind. Ein Beispiel hierfür kann die Komplexität der Aufgabe sein. Durch die kapazitive Beschränktheit der mentalen Ressourcen des Menschen (vgl. Kap. 1.2.1.1) kommt es vor, dass bei der Ausführung komplexer Aufgaben mehr Fehler gemacht werden. In Verbindung mit der ungenauen Erinnerung aus dem Langzeitgedächtnis kommt es dazu, dass komplexe Folgen von Aufgabenelementen zu einer Überbeanspruchung des menschlichen Gedächtnisses und daraufhin zu Fehlern führen [301]. In diesem Fall spricht man von der Aufgabenkomplexität als Ursache des Fehlers. Wenn äußere Umstände die Ursache für menschliche Fehler sind, dann spricht man von einer fehlerbegünstigenden Situation. Beispiele hierfür sind: schlechte Arbeitsplatzgestaltung, besondere Umwelteinflüsse, Stress (vgl. Kap. 2.2.1) unangemessene Lern- und Arbeitshilfen oder eine nicht ausreichende Unterweisung [263]. Zusätzlich können auch Leistungsvoraussetzungen Ursachen für menschliche Fehler sein. Hierzu zählen neben mangelnden Fähigkeiten, Fertigkeiten und ungenügendem Wissen auch Motivation und Einstellung [301].

Im medizinischen Arbeitssystem (Kap. 1.2.2.1) kommen zu den inneren und äußeren Einflussfaktoren noch patientenabhängige Faktoren hinzu, wie z.B. anatomische Besonderheiten und nicht diagnostizierte Begleiterkrankungen [388]. Es ist offensichtlich, dass der menschliche Fehler seinen Ursprung nicht nur im menschlichen Versagen hat, sondern auch durch viele zusätzliche Faktoren beeinflusst wird. Darum kann auch in den meisten Fällen nicht der Anwender alleine für den Fehler verantwortlich gemacht werden.

### **5.1.2 Klassifizierungen menschlicher Fehler**

*“To err is human...”*

Cicero

*“...to understand the reasons why humans err is science”*

E. Hollnagel

In der Literatur lassen sich eine Vielzahl von Kategorienschemata für menschliche Fehler, Fehlhandlungen und Handlungsfehler finden [293, 294, 301, 325, 327,

332, 432]. Für die in dieser Arbeit ergründete Thematik erwies sich das von Reason entwickelte Grundschema zur Fehlerbeschreibung als geeigneter Ausgangspunkt [329, 330, 332]. Er teilt unter dem Begriff „Fehlverhalten“<sup>43</sup> die menschlichen Fehler in die Oberkategorien: 1. „Schnitzer/ Patzer“<sup>44</sup> (manchmal auch slip = Ausrutscher und lapse = Aussetzer genannt) als Fehlverhalten im Stadium der Ausführung und 2. „Irrtum“<sup>45</sup> als Fehlverhalten in den Stadien der Beurteilung, der Schlussfolgerung und der Planung ein [329, 332, 429]. Reason klassifiziert Fehlertypen mit dem Generischen Fehler-Modellierungs-System (GFMS)<sup>46</sup> [331]. Aufbauend auf den Verhaltenskategorien von Rasmussen [326] bezieht er die Fehlermechanismen auf die drei bereits in Kapitel 1.2.1.1 beschriebenen unterschiedlichen Ausführungsebenen (Abb. 64) [331, 332]. „Patzer und Schnitzer“ werden der fähigkeitsbasierten Ebene zugeordnet, wobei sie der Entdeckung eines Problems vorausgehen und unter dem Einfluss spezieller Schemata stehen. Sie sind in der Regel Folgen nicht ausreichender Aufmerksamkeitsüberwachung. Dagegen werden Irrtümer der regelbasierten und wissensbasierten Ebene zugeordnet, da sie im Normalfall zeitlich nach Auftreten eines Problems angesiedelt sind, wobei die regelbasierte Ebene an bestimmte Bedingungen gekoppelt ist. Probleme werden in dieser Ebene durch typische Wenn-dann-Beziehungen gelöst. Fehler sind meist missglückte Problemlösungsversuche und entstehen durch Fehleinschätzungen. Hieraus folgen falsche Anwendungen der Regeln, sowie falsche Abrufe der Prozeduren. Die wissensbasierte Ebene ist dagegen mit Bewusstseins-, Arbeitsgedächtnis- und Aufmerksamkeitsprozessen verknüpft. In dieser Ebene treten für den Menschen neue Situationen auf, welche durch analytische Prozesse und gespeichertes Wissen bewältigt werden müssen. Fehler entstehen durch beschränkte Ressourcen und durch unvollständiges oder fehlerhaftes Wissen [331, 332, 354]. Nach Erkenntnissen von Reason sind Menschen am anfälligsten gegenüber „Patzen und Schnitzern“. 60,7% der Fehler treten in

---

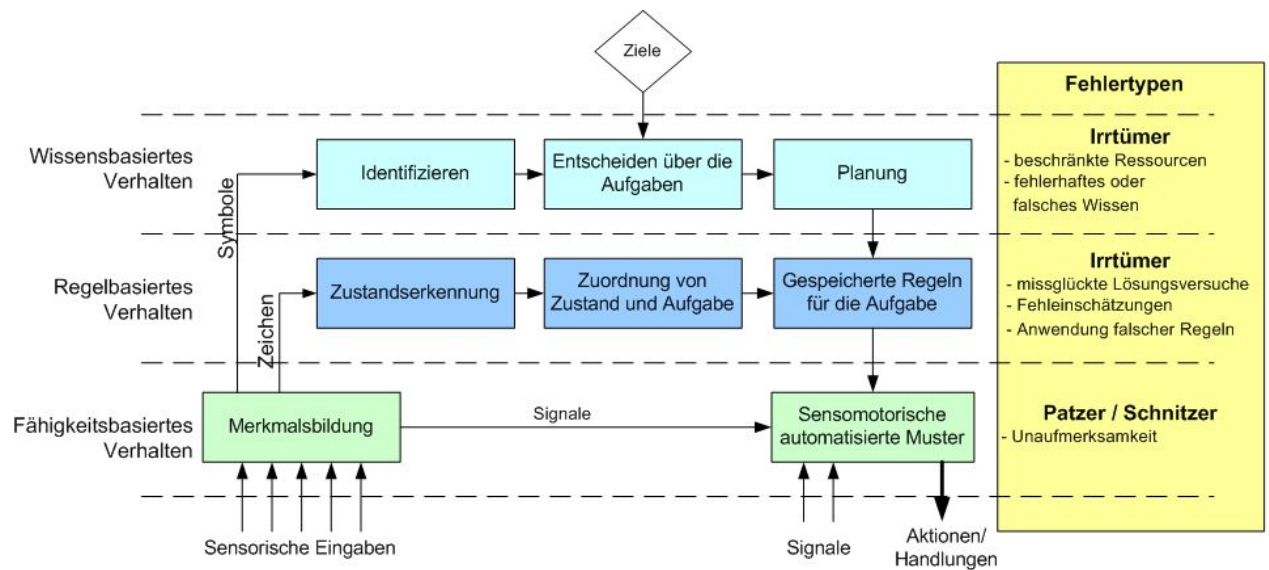
<sup>43</sup> Reason definiert Fehlverhalten als Oberbegriff, „der all die Ereignisse umfasst, bei denen eine geplante Abfolge geistiger oder körperlicher Tätigkeiten nicht zum beabsichtigten Resultat führt, sofern diese Misserfolge nicht fremdem Einwirken zugeschrieben werden können“ [330].

<sup>44</sup> engl.: slips and lapses - Reason: “Slips and lapses are errors which result from some failure in the execution and/or storage stage of an action sequence, regardless of whether or not the plan which guided them was adequate to achieve its objective” [332].

<sup>45</sup> engl.: mistakes - “Mistakes may be defined as deficiencies or failures in the judgemental and/or inferential processes involved in the selection of an objective or in the specification of the means to achieve it, irrespective of whether or not the actions directed by this decision-scheme run according to plan” [332].

<sup>46</sup> engl. Generic error-modelling system (GEMS)

der fähigkeitsbasierten Ebene auf, 27,1% in der regelbasierten und 11,3% in der wissensbasierten Ebene [332].



**Abb. 64: Menschliche Verarbeitungsebenen mit typischen Fehlern [326, 331]**

Die gleiche, von Reason angewandte Terminologie der Fehlerunterteilung mit Fehlern in der Absicht (mistake) und Fehlern bei der Ausführung (slips) verwendet auch Norman [294, 314]. Beide Fehlerklassen hält er für sehr wichtig, da jede für sich unterschiedliche Formen der Problematik verdeutlicht, jeweils mit unterschiedlichen Grundregeln und Lösungsvorschlägen. Slips klassifiziert er in Unterklassen nach der scheinbar möglichsten Ursache (Tab. 21).

Haupt-kategorie	Unterklasse	Grund
Fehler in der Zielbildung	Wissensbasierte Fehler	Irrtümer, mangelnde Kenntnis
	Verwechslungsfehler	Falsche Klassifikation
	Beschreibungsfehler	Verwechslung von Bewegungsprogrammen durch ähnliche Situationsmerkmale
Aktivierungsfehler	Unbeabsichtigte Aktivierung	
	a) Stereotypisierungsfehler	Bei ähnlichen Situationsmerkmalen gewinnt das geläufigere Merkmal die Oberhand im Handlungsablauf.
	b) Externe Aktivierung	Änderung der Handlungsausführung durch äußere Situationsmerkmale
	c) Assoziationsmerkmale	Verknüpfte Schemata werden gemeinsam realisiert, auch wenn nur eins Sinn macht.
	Aktivierungsverlust	Vergessen von Handlungszielen oder Auslassen von Schritten in der Handlungsabfolge.
Falscher Aufruf aktiver Schemata	Falsche Auslösebedingung	Falscher Zeitpunkt der Aktivierung
	Fehlende Auslösebedingung	Schema wird nicht aufgerufen.

**Tab. 21: Klassifizierung der Slips (Fehler) nach vermuteten Ursachen [293, 294 mit Übersetzung von 118]**



Die grundlegende Klassifikation basiert darauf, dass jede mögliche Absicht eines Menschen eine Anzahl Schemata aktiviert und dann eine Handlungssequenz kontrolliert. Diese Schemata sind organisierte Wissensstrukturen, welche jeweils mit einem Aktivierungswert und einem eigenen Satz von Auslösebedingungen verknüpft sind [293]. Übereinstimmend mit Reason macht Norman die Beobachtung, dass Handlungsfehler an Entscheidungspunkten auftreten [293, 332].

Im Gegensatz zu Reasons Fehlerklassifikation [332] stützt sich Norman auf die bereits in Kapitel 1.2.1.1 beschriebene Aktionstheorie (Activation-Trigger-System) [293, 354, 295]. Er orientiert sich an Fehlern während der Entwicklung der Absicht, fehlerhaften Aktivierungen von Schemata<sup>47</sup> und Fehlern durch falsche Auslösung der Schemata (Tab. 21) [293]. Durch die Betrachtung von Feedback-Mechanismen beabsichtigter Aktionen eignet sich Normans Klassifizierung für interaktive Geräte etwas besser als die von Reason verwendete [180, 294]. Trotzdem bleibt bei dieser Fehlerklassifizierungsmethode für den Kontext der Mensch-Maschine-Kommunikation die Problematik, dass sie ausschließlich auf den Menschen als Ursache des Fehlers fokussiert. Gleichwohl kann bei einer Mensch-Maschine-Kommunikation die Fehlerursache zusätzlich im Kommunikationsversagen zwischen Mensch und Maschine oder auch nur bei der Maschine liegen. Schnell wird bei der Fehlerbetrachtung vergessen, dass der Anwender u.U. durch die Konstruktion des technischen Systems in eine kritische oder fehlerträchtige Situation gebracht wird [354].

Rasmussens Modell der Aktionsregulation besteht aus fünf mentalen Funktionen: Beobachten, Identifizieren, Interpretieren, Handlungen definieren und Handlungen ausführen. Er klassifiziert Fehler anhand dieser mentalen Funktionen sowie der bei Norman beschriebenen menschlichen Verarbeitungsebenen [325]. Tab. 22 veranschaulicht ein Fehlerklassifikationsschema, aufbauend auf der Theorie nach Rasmussen.

Weitere Klassifizierungsmöglichkeiten gliedern sich nach der Fehlerart in zufällige-, systematische oder beabsichtigte Fehler, nach der Fehlerdauer in permanente, periodische und temporäre Fehler und nach dem Fehlerursprung in menschliche und physikalische Fehler bzw. in Fehler, welche innerhalb der Systemgrenzen liegen.

---

<sup>47</sup> organisierte Wissensstrukturen, welche Handlungssequenzen kontrollieren [293]

Mentale Funktionen	Verarbeitungsebenen		
	Fähigkeitsbasiert	Regelbasiert	Wissensbasiert
Beobachten	Inadäquate Musterentwicklung Wahrnehmungsfehler	Beobachtungsfehler	Lesefehler: Informationsmangel
Identifizieren	Merkmalsübereinstimmungsfehler	Prüffehler Klassifizierungsfehler	Diagnosefehler
Interpretieren			Prognosefehler - Zielfehler - falsche Hypothese Evaluationsfehler - falsches Ergebnis - falsche Akzeptanz - falsche Ablehnung
Handlungen definieren		Auswahlfehler	Entscheidungsfehler
Handlungen ausführen	Prozedurfehler Durchführungsfehler Ausrutscher	Abruffehler Ablauffehler	Planungsfehler

Tab. 22: Fehlerklassifizierungsschema nach Rasmussen [8, 325]

### 5.1.3 Fehler der Mensch-Maschine-Kommunikation und Klassifizierung von Gebrauchstauglichkeitsproblemen

Wie bereits deutlich wurde, muss eine falsche oder fehlerhafte Geräteverwendung nicht unbedingt auf einen Fehler des Menschen zurückzuführen sein. Derartige Fehler können viele Faktoren als Ursache haben, welche auf Probleme in der Mensch-Maschine-Kommunikation zurückzuführen sind. Gaba et al. sprechen im Fall des Mensch-Maschine-Versagens von latenten Fehlern, die sich beispielsweise in einem unzureichenden Gerätedesign bzw. einer fehlenden Unterstützung des Anwenders äußern [144]. Es wird somit von einem Fehler gesprochen, der auf mangelnde Gebrauchstauglichkeit (vgl. Kap. 2) des Gerätes zurückzuführen ist. Krushniruk et al. zeigen einen direkten Zusammenhang zwischen Gebrauchstauglichkeitsproblemen und Fehlern (slips & mistakes) bei medizinischen Geräten [226].

Noch vor nicht allzu langer Zeit war es generell üblich, dass bei Verletzungen eines Patienten durch falsche Gerätebedienung oder anormalen Gebrauch der Hersteller die Schuld auf den unerfahrenen Nutzer abwälzen konnte [346]. Diese Denkweise hat sich jedoch in den USA inzwischen stark und in Deutschland zumindest ein wenig verändert. In den USA können Hersteller durch die FDA bei benutzerbezogenen Vorfällen (engl. „user-related incidents“) zur Verantwortung gezogen werden. Hierzu werden neuerdings neben den Anwenderfehlern („user

error“) auch die Anwendungsfehler (engl. „use error“) betrachtet [310]. In Deutschland erfolgt dieser Druck auf die Hersteller durch die in Kapitel 2.3 beschriebenen Normen DIN EN IEC 60601-1-6 und DIN EN IEC 62366, welche davon ausgehen, dass gerätebezogene Fehler gewöhnlich aus vielfachen, miteinander zusammenhängenden Faktoren resultieren. Schadensberichte bestätigen, dass das Gerät oft selbst zu Gebrauchsfehlern beiträgt, was in aller Regel durch Konzeptionsfehler der Schnittstelle zwischen Anwender und Gerät verursacht wird [97].

Oft wird fälschlicherweise für derartige Fehler der Begriff „Usability Fehler“ oder Gebrauchstauglichkeitsfehler verwendet. Gemeint ist jedoch ein Fehler, welcher durch ein Gebrauchstauglichkeitsproblem verursacht oder verstärkt wurde. Die DIN EN IEC 60601-1-6 spricht hierbei von einem Gebrauchsfehler<sup>48</sup> [97], welcher mangelnde Gebrauchstauglichkeit des medizinischen Gerätes als Ursache hat. Meist handelt es sich hierbei um Fehler die Interaktions- und Kommunikationsprobleme zwischen dem Anwender und dem System als Ursache haben.

In der Literatur gibt es derzeit keine eindeutige Definition für ein Gebrauchstauglichkeitsproblem. Wenn man von der in der DIN EN 9241-11 beschriebenen Definition von Gebrauchstauglichkeit als „das Ausmaß, in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufrieden stellend zu erreichen“ ausgeht [113], müsste ein Gebrauchstauglichkeitsproblem vorliegen, sobald ein Nutzer sein Ziel nicht unter diesen Kriterien erreicht. Dann müsste allerdings jeder Maschinen- und menschliche Fehler ein Gebrauchstauglichkeitsproblem darstellen. Im eigentlichen Sinne ist daher ein Gebrauchstauglichkeitsproblem ein Problem an der Mensch-Maschine-Schnittstelle zwischen Anwender und Gerät. Generell ist ein Gebrauchstauglichkeitsproblem ein Umstand, der die Gebrauchstauglichkeit beeinträchtigt. Nach Karat et al. ist ein Gebrauchstauglichkeitsproblem „alles, was mit der Fähigkeit des Nutzers interferiert, seine Aufgabe effizient und effektiv zu komplettieren“ [210]. Somit muss ein Gebrauchstauglichkeitsproblem nicht nur durch das System hervorgerufen werden, sondern kann auch durch die Umgebung oder die Arbeitsabläufe entstehen. Entsprechend der DIN EN 9241-11 gilt Gebrauchstauglichkeit für einen bestimmten Nutzer in einem bestimmten Nutzungskontext. Somit könnte eine Definition lauten:

---

<sup>48</sup> Handlung oder Unterlassung einer Handlung, die ein anderes Ergebnis hat als vom Hersteller beabsichtigt oder vom Anwender erwartet. Gebrauchsfehler beinhalten Versehen, Vergessen, Fehler und vernünftigerweise vorhersehbaren Missbrauch.

*„Ein Gebrauchstauglichkeitsproblem liegt dann vor, wenn die Mensch-Maschine-Interaktion gestört wird und ein bestimmter Benutzer in einem bestimmten Nutzungskontext sein Ziel nicht effektiv, effizient und zufrieden stellend erreichen kann.“*

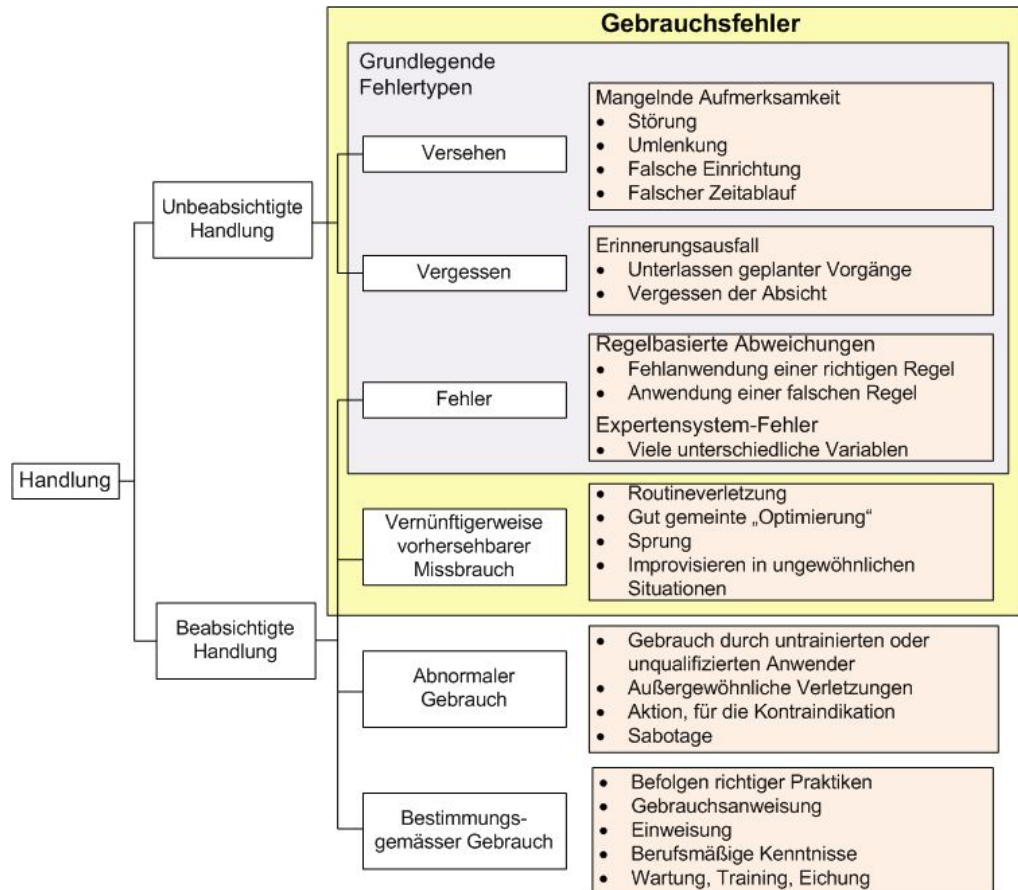
Beispiele für derartige Probleme sind in Tab. 23 dargestellt.

Gebrauchstauglichkeitsproblem	Möglicher resultierender Gebrauchsfehler
Druckknöpfe auf einem Bedienfeld sind zu eng nebeneinander angebracht.	Der Anwender drückt den falschen Knopf.
Zwei Symbole auf einem Bildschirm sehen sehr ähnlich aus.	Der Anwender legt das Symbol falsch aus und wählt die falsche Funktion.
Eine Schnittstelle zwischen Anwender und Gerät erfordert eine komplexe, langwierige und willkürliche Reihenfolge von Tastendrücken, um eine Infusion zu starten.	Der Anwender gibt die falsche Reihenfolge ein und versäumt damit das Einleiten der Infusion.
Vom Anwender eingestellte Hoch-/Niedrig-Alarmgrenzen eines Herzfrequenzmonitors werden nicht kontinuierlich angezeigt.	Der Anwender versäumt es, einen gefährlichen Anstieg der Herzfrequenz zu entdecken, weil die Alarmgrenzen zu hoch eingestellt sind und der Anwender sich zu sehr auf das Alarmsystem verlässt.
Die vom Anwender normalerweise angewandte Kraft übersteigt die Bruchfestigkeit eines Katheters.	Der Anwender zerbricht den Katheter beim Befestigen.

**Tab. 23: Beispiele von Gebrauchstauglichkeitsproblemen und daraus resultierenden Fehlern [97]**

In der Literatur werden verschiedene Klassifikationssysteme für Fehler, welche auf Gebrauchstauglichkeitsprobleme zurückzuführen sind, sowie für Gebrauchstauglichkeitsprobleme vorgeschlagen. Der Standard ANSI-NCITS 354 [5] rät auch zu der Verwendung eines Klassifizierungssystems und nennt als Beispiel das von Norman entwickelte (vgl. Kap. 5.1.2). Dieses Klassifizierungssystem ist jedoch hauptsächlich für menschliche Fehler konzipiert worden und demnach zu wenig auf die Interaktivität zwischen Mensch und Maschine ausgerichtet.

Eine für die Gebrauchstauglichkeitsevaluation geeignete Klassifizierung verwendet die DIN EN IEC 60601-1-6 [97]. Sie unterscheidet Gebrauchsfehler bei unbeabsichtigten Handlungen durch die Klassen Versehen und Vergessen und bei beabsichtigten Handlungen durch die Klassen Fehler und den vernünftigerweise vorhersehbaren Missbrauch. Auswirkungen durch anormalen Gebrauch sind nicht der Kategorie Gebrauchsfehler zugeordnet, da sie beispielsweise durch Sabotage oder durch nicht qualifizierte Anwender verursacht werden (Abb. 65).

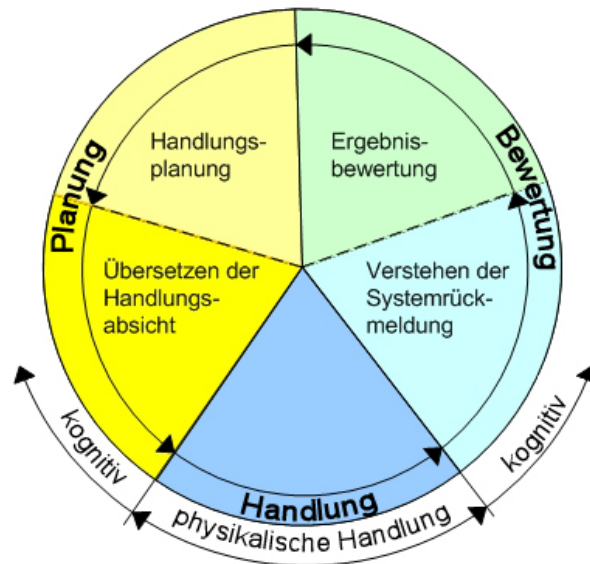


**Abb. 65: Klassifikation von Handlungen adaptiert nach der DIN EN IEC 60601-1-6 [97]**

Weitere Klassifizierungssysteme lassen sich u.a. bei Booth [34], Fu et al. [142], Havenberg & Law [195] sowie Frederics & Basili [141] finden. Aus dieser Vielfalt der Klassifizierungssysteme erkennt man die Uneinheitlichkeit dieses Problems. Um der in Kapitel 4.1.8 geforderten und beschriebenen Reliabilität gerecht zu werden, muss ein einheitliches Klassifizierungssystem für Nutzungs-, bzw. Gebrauchstauglichkeitsprobleme verwendet werden.

Laut einer Studie von Hamborg et al. [165] erfüllt das von Andre et al. vorgestellte User Action Framework (UAF) [3] diese Forderung im Sinne einer reliablen Beschreibung und Kategorisierung von Gebrauchstauglichkeitsproblemen. Aufbauend auf dem durch Norman entwickelten Handlungsmodell [295] und dem durch Keenan et al. beschriebenen Gebrauchstauglichkeitsproblem Taxonomie [213], unterteilt das UAF die Gebrauchstauglichkeitsprobleme in sieben verschiedene Handlungsphasen (vgl. Kap. 1.2.1.1), welche in der ersten Ebene den Phasen Planung, Handlungsausführung und Bewertung zugeordnet sind (Abb. 66). In einem ersten Schritt werden die Gebrauchstauglichkeitsprobleme diesen drei Kategorien zugeordnet, um dann sukzessive über weitere Hierarchieebenen in

Bezug auf handlungs- und gestaltungsbezogene Ursachen konkretisiert zu werden [3, 165, 193]. Eine annähernd eindeutige Eingrenzung der Ursache von Gebrauchstauglichkeitsproblemen ist durch diese schrittweise Konkretisierung möglich. Der Anwender wird hierbei durch ein Software-Tool, den UAF-Viewer<sup>49</sup>, unterstützt.



Normans Handlungsmodell	Phase im Interaktionszyklus	Bewertungsaspekte
Festlegung des Handlungsziels	Planung – hohe Ebene	Können Nutzer die allgemeinen Anforderungen bestimmen, um mit einer Handlung beginnen zu können?
Bildung der Handlungsabsicht	Planung – hohe Ebene	Können Nutzer festlegen, was zu tun ist, um eine Handlungsabsicht zu bilden?
Handlungsspezifikation	Planung – Übersetzung der Handlungsabsicht	Können Nutzer festlegen, wie sie ihre Handlungsabsicht in Handlungen umsetzen?
Handlungsausführung	Physische Handlungsausführung	Können Nutzer Handlungen einfach ausführen?
Wahrnehmung des Systemzustandes	Bewertung – Systemrückmeldung wahrnehmen	Können Nutzer die Systemrückmeldung wahrnehmen?
Interpretation des Systemzustandes	Bewertung – Systemrückmeldung verstehen	Können Nutzer die Systemrückmeldung verstehen?
Evaluation des Systemzustandes mit Bezug auf Ziele und Intentionen	Bewertung – Bewertung des Ergebnisses	Können Nutzer den Erfolg des Ergebnisses feststellen?

**Abb. 66: UAF-Interaktionszyklus und Zusammenhang zwischen Normans Handlungsmodell und dem UAF-Interaktionszyklus [3, 165]**

<sup>49</sup> UAF-Viewer: englisches Original unter: <http://hemlock.cs.vt.edu/uaf>

#### **5.1.4 Klassifizierung von Fehlern in sicherheitskritischen Systemen**

Bei Störfällen und Anomalien in sicherheitskritischen Systemen werden oft menschliches und technisches Versagen unterschieden. Herczeg erklärt, dass eine solche Klassifizierung nicht angemessen ist, und daraus falsche Schlussfolgerungen gezogen werden. Er fügt dieser Unterscheidung noch die Gruppe der Interaktions- und Kommunikationsfehler hinzu, welche aus einem Interaktions- und Kommunikationsversagen des Mensch-Maschine-Systems resultieren [180]. In sicherheitskritischen Systemen (► Kap. 5.3.1) kann ein Unfall das Ergebnis vieler Fehler sein, die z.B. durch das Design, das Equipment, die Abläufe, die Operatoren oder die Umgebung hervorgerufen werden [308]. Zu dieser Erkenntnis ist Rasmussen bereits im Jahr 1982 gekommen. Er sah nicht allein den Menschen, resp. menschliche Fehler, als Ursachen für jegliche Art technischer Störungen an. Vielmehr hielt er es für richtig, menschliche Fehler als Instanzen von Mensch-Maschine- oder Mensch-Aufgaben-Problemen zu betrachten [325]. Seiner Ansicht nach sollte im Fall von systematischen Fehlern auf einen Designfehler geschlossen werden. Andererseits sind gelegentliche Fehler normalerweise auf Schwankungen in Teilen des Systems oder auf den Menschen zurückzuführen und werden als menschliche Fehler und Systemfehler betrachtet.

Gerade in sicherheitskritischen Systemen ist es notwendig, die Merkmale einer fehlerhaften Aufgabe zu identifizieren und zu charakterisieren, um die Ursache des Fehlers beseitigen zu können. Rasmussen beschreibt hierzu die Ursachen, Mechanismen und Auswirkungen menschlicher Fehler, welche miteinander korrelieren können, unter Berücksichtigung äußerer Einflüsse bzw. emotionaler und motivierender Aspekte der Arbeitssituation sowie physiologischer Faktoren. Er stellt sie in einem Taxonomieschema menschlicher Fehler dar (Abb. 67).

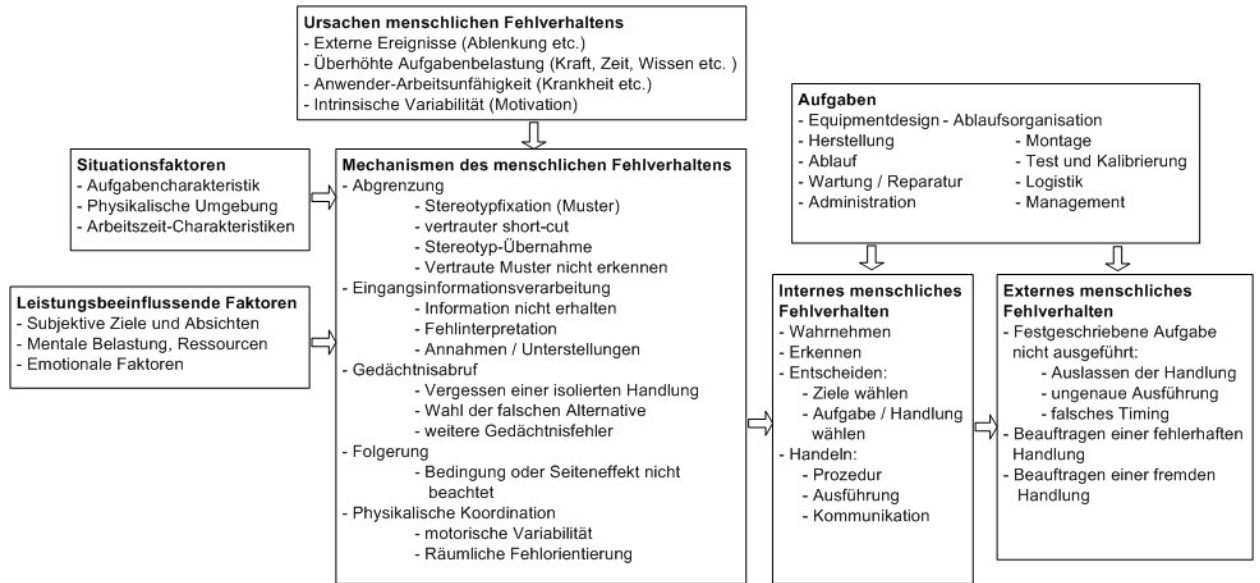


Abb. 67: Ursachen, Mechanismen und Auswirkungen menschlicher Fehler nach Rasmussen [325]

Fehler, welche in den verschiedenen Phasen der Interaktion auftreten können, haben Galliers et al. in Normans Interaktionsmodell (vgl. Kap. 1.2.1.1) eingebunden (Abb. 68) und entsprechende Anforderungen an sicherheitskritische Geräte aufgestellt, um diese Fehler zu verhindern oder die Auswirkungen zu verringern (Tab. 24).

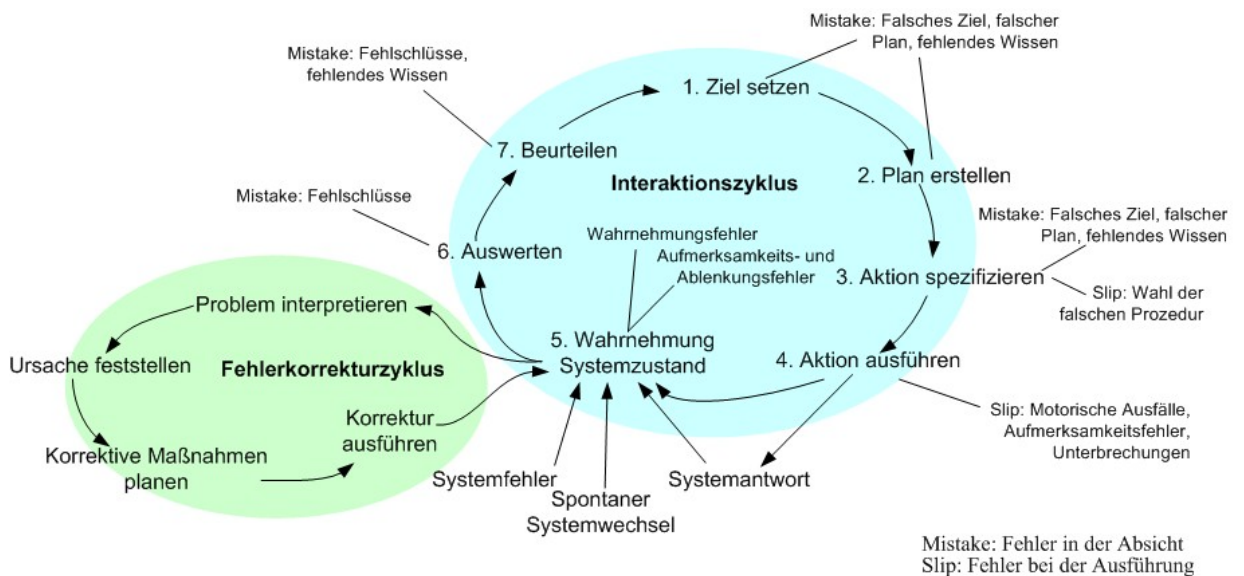


Abb. 68: Fehler im Interaktionszyklus [146, 295]



Phase im Interaktionszyklus	Ausrutscher (slip) Gegenmaßnahme	Irrtum (mistake) Gegenmaßnahme
Ziel formulieren	-	Metaphern verwenden, Systemmodelle, Hinweise verwenden
Plan erstellen	-	Unsichere Handlungen blockieren, Gedächtnisstützen, Systemsimulation, funktionsfähige Metaphern, eindeutige Bedienelemente
Aktion spezifizieren	Bei schwierigen Handlungen sichere Handlungen vorauswählen	Direkte Hinweise, Geräteanforderungscharakter
Aktion ausführen	Sicherheitsschleusen, einfache Bedienung, Störungssicherheit, Gefährliche Situationen vermeiden, Eingabeprüfung	Eingabeprüfung
Systemzustand wahrnehmen	Herausstechende Warnung, Multimodale Warnung, klare Meldungen	-
Auswerten	Klare Meldungen, wiederkehrende Warnmeldungen	Unmissverständliches Feedback
Beurteilen	-	Systemmodelle, um Änderungen und Änderungsfolgen auszugeben, Modelle zum Situationsbewusstsein

**Tab. 24: Fehlergegenmaßnahmen in den einzelnen Phasen der Interaktion [146]**

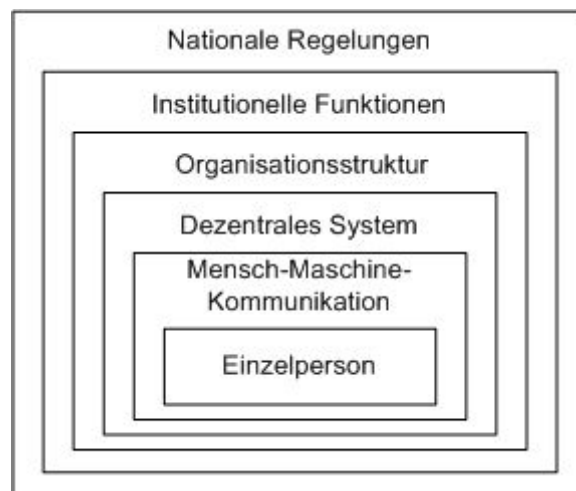
### 5.1.5 Klassifikation von Fehlern in der Medizin

Leider kann sich keine Disziplin von Fehlern freisprechen, also auch nicht die Medizin. Gerade in diesem Bereich sind Fehler allerdings oft schwerwiegend, sie können gesundheitsschädigend sein oder sogar tödlich enden. Weiterhin führen sie zu erheblichen wirtschaftlichen Folgen [157, 220].

Wie bereits in Kapitel 5.1.2 beschrieben, ist nach Reason ein menschlicher Fehler ein Fehler in der menschlichen Handlung [220]. Ausgehend davon, dass die menschliche Handlung in erster Linie eine kognitive Aktivität ist, spielen auch bei medizinischen Fehlern kognitive Faktoren eine fundamentale Rolle. Dies gilt nicht nur für die einzelnen Individuen, sondern lässt sich in mehreren Ebenen auf die Gesamtheit des Gesundheitswesens übertragen (Abb. 69).

Auf der ersten Ebene der Fehlerursachen stehen Einzelpersonen. Kognitive Faktoren spielen hier die größte Rolle [332]. Eine Instanz höher können Fehler bei der Interaktion einer Einzelperson mit der Technik auftreten. Dies ist Teil der Mensch-Maschine-Kommunikation, bei der kognitives Verhalten die Interaktion zwischen Mensch und Technik beeinflusst und manchmal das menschliche Handeln bestimmt [176, 427, 429]. Auf der dritten Ebene können Fehler der sozialen Dynamik in der Interaktion zwischen Personengruppen zugeordnet werden, welche mit komplexer Technologie in einem dezentralen kognitiven System interagie-

ren. In den Ebenen vier bis sechs können Fehler den Faktoren der Organisationsstruktur (z.B. Koordination, Kommunikation, Normierung des Arbeitsprozesses), den institutionellen Funktionen (z.B. Regelungen und Richtlinien) und nationalen Regelungen zugeschrieben werden. Auch auf diesen höheren Ebenen spielen kognitive Faktoren eine Rolle. Obwohl die Fehler auf den sechs Ebenen unabhängig voneinander untersucht werden können, ist eine Betrachtung des gesamten Systems für ein eingehendes Verständnis der medizinischen Fehler wichtig [429].



**Abb. 69: Systemhierarchie menschlicher Fehler in der Medizin [429]**

Für einen einheitlichen Fachwortschatz innerhalb des Bereichs „Fehler in der Medizin“ hat ein 16-köpfiger Expertenkreis eine Nomenklatur und Definitionen erarbeitet und in einem Glossar veröffentlicht [388]. Generell bietet dieses Werk zwar keine neuen Erkenntnisse, aber eine gute Zusammenfassung der bisherigen und ihre Überführung in den medizinischen Bereich. Dieser Nomenklatur nach ist ein medizinischer Fehler „jede geplante Vorgehensweise, die nicht plangemäß ausgeführt wurde (Ausführungsfehler), oder das Anwenden einer Vorgehensweise, die zum Erreichen eines Ziels ungeeignet ist (Planungsfehler)“ [388]. Laut Glossar wird in Deutschland nur zwischen groben Behandlungsfehlern bzw. vermeidbaren und unvermeidbaren Behandlungsfehlern unterschieden [388]. Auf internationaler Ebene lassen sich jedoch weitere Klassifikationen finden. Die Joint Commission on Accreditation of Healthcare Organizations (JCAHO) hat die gebräuchlichsten Klassifikationen für medizinische Fehler zusammengestellt, deren Abhängigkeiten in Abb. 70 dargestellt sind [59]:

1. Die Klassifikation eines medizinischen Fehlers nach der Auswirkung unterscheidet zwischen medizinischen und nicht medizinischen Auswirkungen. Medizinische Auswirkungen werden in erster Ebene in psychische und physische, nicht medizinische in juristische, soziale und ökonomische Auswirkungen unterteilt.
2. Die Klassifikation nach dem Fehlertyp unterscheidet die drei Hauptebenen Kommunikation, Patienten-Management und klinische Leistungsstärke mit den jeweiligen Unterebenen. Die Kommunikations-Unterebenen beschreiben Kommunikationsprobleme zwischen medizinischem Personal und Patienten, Patientenvertretern und nicht medizinischem Personal, bzw. auch medizinischem Personal untereinander. Nicht unterschieden werden die für diese Arbeit relevanten Kommunikationsprobleme zwischen medizinischem Personal und dem technischen System.
3. Die Klassifizierung nach dem Arbeitsbereich unterscheidet die Gegebenheiten der Arbeitsumgebung, in der ein Fehler auftritt, sowie die Personen, welche involviert sind.
4. Die Klassifizierung eines medizinischen Fehlers nach der Ursache ist unterteilt in System- und menschliche Fehler. Nach dieser Klassifikation sind Systemfehler organisations- oder technisch bedingt. Bei technischen Fehlern sind auch die für diese Arbeit interessanten Fehler, welche auf das Gerätedesign zurückzuführen sind, eingegliedert. Menschliche Fehler werden weiterhin in Fehler unterteilt, die durch den Patienten, das Personal oder externe Faktoren hervorgerufen werden. Fehler des Personals sind weiterhin nach der durch Rasmussen (vgl. Kap. 5.1.2) bekannten Klassifikation in fähigkeitsbasierte, regelbasierte und wissensbasierte Fehler aufgeschlüsselt.

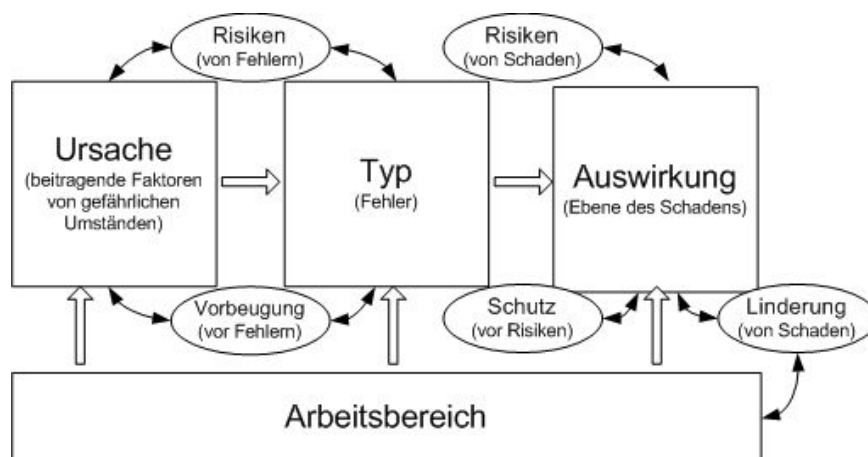


Abb. 70: Zusammenwirken der Klassifizierungen medizinischer Fehler nach der JCAHO [59]

Für die in dieser Arbeit relevante Fragestellung ist hauptsächlich eine Klassifikation nach der Fehlerursache von Interesse. Die Klassifikationen und Taxonomien, wie sie in dem „Glossar für Patientensicherheit“ und bei der JCAHO zu finden sind, liefern eine gute allgemeine Taxonomie und Klassifizierungsmöglichkeit medizinischer Fehler. Auch werden, wenn auch nur am Rande, Fehler des technischen Systems betrachtet. In dem Glossar für Patientensicherheit fallen diese in die Kategorie der „latenten Fehler“<sup>50</sup> und in der Klassifikation der JCAHO in den Bereich, der als Ursache direkt das technische System sieht. Diese nur oberflächliche Betrachtung des technischen Systems ist jedoch für die Zielstellung dieser Arbeit nicht ausreichend. Auch bleiben Fehler in der Kommunikation zwischen dem Menschen und dem technischen System außer Acht.

Aufbauend auf Reason's Definition von Fehlern [332] (vgl. Kap 5.1.2) und Norman's kognitiver Theorie der menschlichen Handlung (vgl. Kap 1.2.1.1), haben Zhang et al. eine handlungsbezogene, kognitive Fehlertaxonomie entwickelt, welche weitgehend folgenden Kriterien im Bereich der Medizin gerecht wird [429]:

1. Kategorisierung aller Fehlertypen nach einer kognitiven Dimension,
2. Zuordnung eines jeden Fehlers zu einem speziell zugrunde liegenden kognitiven Mechanismus,
3. Erklärung des Grundes bzw. Vorhersagung von Zeit und Ort des Auftretens spezieller Fehler,
4. Entwicklung von Interventionsstrategien für jeden Fehlertyp.

In dieser Fehlertaxonomie wird auf das bereits von Norman und Reason verwendete Vokabular zurückgegriffen. Versehen und Irrtümer werden in der ersten Ebene in Ausführungsversehen und Bewertungsversehen, sowie in Zielirrtum, Absichtsirrtum und Handlungsirrtum unterteilt. Eine weitere Unterteilung mit entsprechenden Praxisbeispielen findet sich in Tab. 25.

---

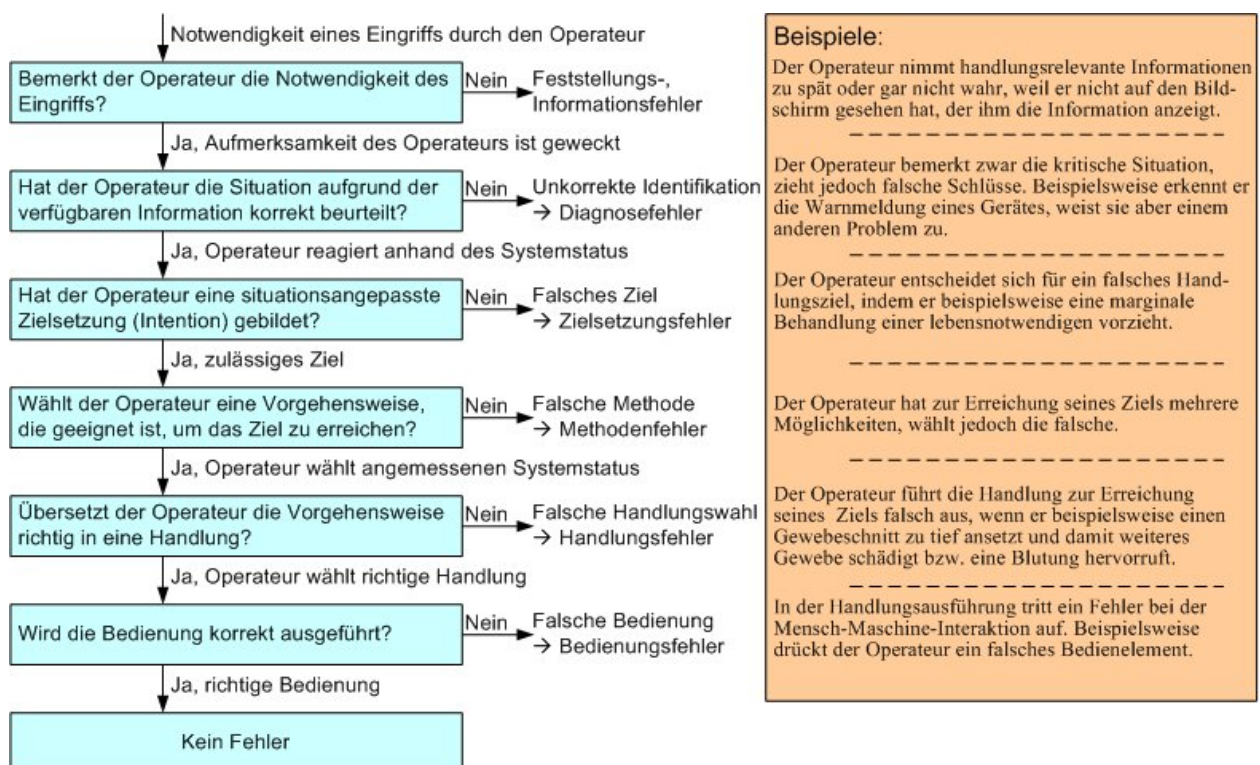
<sup>50</sup> Latente Fehler sind Mängel im System, wie etwa schlechtes Design, falsche Installation, ungenügende Wartung, ungünstige Einkaufsentscheidungen und unzureichende Personalausstattung [388].

Phase im Aktionszyklus		Beispiel	Kognitiver Mechanismus	Potentielle Lösung
Versehen bei der Ausführung	Versehen bei der Zieldefinition	Ein Arzt wird für einen wichtigen Telefonanruf aus dem Patientenzimmer gerufen. Danach geht er bereits zu dem nächsten Patienten, ohne die Visite beim vorherigen zu beenden (Aktivierungsverlust).	-Aktivierungsverlust (Vergessen) -Querreden -geänderte Ziele -Verspätete Aktivierung -Überflutung des Zielstapels	Erinnerungshilfen anbieten Multitasking verringern Unterbrechungen verringern Zielstapel / Aufgabenstapel verringern Nutzer trainieren
	Versehen bei der Absicht	Man geht zu einem bestimmten Ort, mit der Intention, etwas dort zu erledigen. Angekommen an dem Ort, weiß man nicht mehr, was man wollte.	-Aktivierungsverlust (Vergessen) -Querreden (gleichzeitig) -Querreden (der Reihe nach) -Aufhebung von Schemata -Aktivierung falscher Schemata	Erinnerungshilfen anbieten Multitasking verringern Unterbrechungen verringern
	Versehen bei der Handlungsfestlegung	Die Arznei IL-11 (Oprelvekin, oder Interleukin-eleven) wurde als IL-2 fehlinterpretiert (Aldesleukin, oder Interleukin-two). 11 wurde als römische zwei gelesen. (Assoziative Aktivierung)	-Assoziative Aktivierung -Fehler in der Abfrage -Sequenzveränderung -Querreden	Automatisierung Entscheidungsunterstützung Nutzertraining Direkte Handlungen
	Versehen bei der Ausführung	„Ich dachte, ich würde nur die „antibiotics IV“ ausstellen, stattdessen habe ich die komplette Infusionspumpe ausgestellt.“ (Capture)	-Wahrnehmungsstörung -Capture Error <sup>51</sup> -Abweichungen bei automatisierten Handlungen -Aussetzer -Versäumnis	Automatisierung Visualisierung Displaydesign Unterbrechungen verringern Erinnerungshilfen
Versehen bei der Beurteilung	Versehen bei der Wahrnehmung	Ein Patient starb an einer flüssigen Aspiration, weil die mit einem Schlauch angeschlossene Flüssigkeitsklappe keinen Mechanismus und kein Systemfeedback hatte, um den Rückfluss in die Trachea zu vermeiden. (Mangel an Wahrnehmung)	-Mangel an Wahrnehmung -Fehlwahrnehmung -Fehlerwartungen	Direkte Wahrnehmung Schnelles Feedback
	Versehen bei der Interpretation	Ein gelbes Blinklicht an einem medizinischen Gerät wurde in einer kritischen Situation als nicht kritisch interpretiert. (Fehlinterpretation)	-Fehlinterpretation -Vorgegebenes Schema -Bestätigungstendenz -Informationsüberlauf -Gedächtnisverlust	-Displaydesign -Entscheidungshilfen -Benutzertraining -Erinnerungshilfen -Situationsbewusstsein
	Versehen bei der Beurteilung	Wegen schlechten Feedbacks wiederholte eine Krankenschwester eine Strahlentherapie dreimal nacheinander. Der Patient starb drei Monate später (Mangel an Feedback)	-Verlorenes Ziel -Mangelhafte Informationen -Beurteilung verschiedener Ziele -Informationsüberlauf -Mangel an Feedback	-Erinnerungshilfen -Displaydesign -Handlungsverfolgung -Reduktion der Information

Tab. 25: Eine handlungsorientierte Taxonomie für Versehen. In Anlehnung an Zhang et al. [429]

<sup>51</sup> „Capture Error: Zwei Aktionssequenzen haben einen ähnlichem Anfang aber ein unterschiedliches Ende. Die besser erlernte Aktionssequenz kann die andere „einfangen“ (Capture), obwohl diese nicht ausgeführt werden sollte. Z.B.: Sonntags zur Arbeit fahren oder einen ausgeliehenen Stift einstecken.

Diese Fehlertaxonomie eignet sich besonders, um menschliche Fehler in der Medizin zu klassifizieren und zu beschreiben. Sie lässt jedoch die Wechselwirkung der Fehler zwischen Mensch und Maschine vermissen. In dem Beispiel „Versehen bei der Handlungsfestlegung“ muss der Fehler nicht unbedingt auf einen rein menschlichen Fehler zurückzuführen sein. Es kann sich auch um einen Fehler in der Kommunikation zwischen Mensch und Maschine handeln. Möglich wäre u.a., dass die Fehlermeldung nicht ausreichend sichtbar ist, sodass sie dem Nutzer verborgen bleibt. Gerade aus diesem Grund ist eine Betrachtung des gesamten Mensch-Maschine-Systems nötig.



**Abb. 71: Das Modell nach Rasmussen [325], adaptiert für den OP-Arbeitsplatz**

Eine weitere Fehlerklassifikationsmöglichkeit für medizinische Fehler ist das Modell der „internalen Fehlfunktion“ nach Rasmussen [325], welches für die Überwachung komplexer technischer Anlagen entwickelt wurde. Mit besonderem Augenmerk auf die Sicherheit betrachtet Rasmussen menschliche Fehler als Instanzen der Mensch-Maschine- oder der Mensch-Aufgaben-Probleme. So kann auch schlechtes Design die Ursache für einen Fehler sein [327]. Rasmussen identifiziert und klassifiziert Fehler anhand von sequentiellen Verknüpfungen zwischen der Informationsaufnahme und der Handlungsausführung. Adaptierte Versionen seines Modells haben sich bereits in verschiedenen sicherheitskriti-

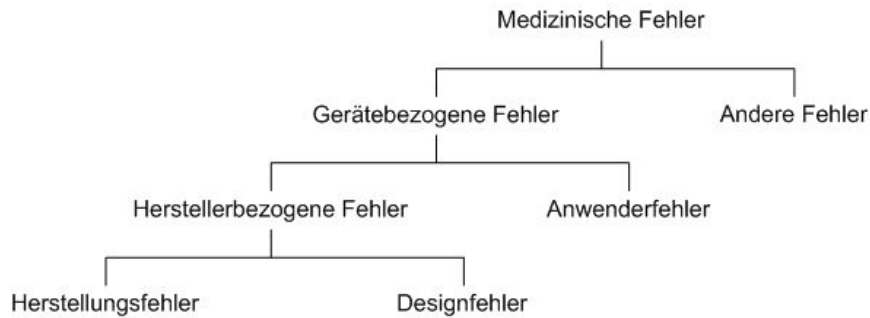
schen Mensch-Maschine-Systemen, wie z.B. in der Luftfahrt bei Flugzeugunglücken [297, 412] oder im Automobilbereich bei Unfällen [431] als geeignet herausgestellt. Abb. 71 zeigt dieses Modell für den OP-Arbeitsplatz adaptiert.

Die Fehlerklassifikation von Rasmussen ist ein geeignetes Instrument, um Fehler in der Umgebung des Operationssaals zu klassifizieren. Ein Vorteil dieser Klassifizierung ist die komplette zeitliche Erfassung von der Informationsaufnahme bis hin zur Handlungsausführung. Zwar bezieht Rasmussen die Klassifizierung hauptsächlich auf den Menschen und betrachtet nicht zusätzlich Fehler der Mensch-Maschine-Schnittstelle, trotzdem können durch diese genaue Bestimmung der Fehlerart Rückschlüsse auf die Verbesserung der Mensch-Maschine-Schnittstelle gezogen werden. Beispielsweise kann ein Informationsfehler Rückschlüsse auf Schwächen der Information durch die Maschine zulassen. Rasmussen klassifiziert eine nicht aufgenommene Information als einen Informationsfehler des Menschen. Bei derartigen Fehlern ist aber auch zu prüfen, ob es sich um ein Versagen der Mensch-Maschine-Schnittstelle handelt. Eine nicht erkannte Warnmeldung an einem Gerät ist daher eher ein Fehler der Mensch-Maschine-Kommunikation, als ein alleiniger Fehler des Operators. So müsste z.B. die Maschine die Warnmeldung mit einem akustischen Signal derart deutlich offerieren, dass der Operator den kritischen Zustand sofort bemerkt, denn schließlich fokussiert er während einer OP eher den Patienten als den Bildschirm des Gerätes.

Gerade an den beiden letzten beschriebenen Klassifizierungen erkennt man die Notwendigkeit der Einbeziehung von Fehlern der Mensch-Maschine-Kommunikation. Eine grobe Einteilung unter der Integration von Gerätefehlern und menschlichen Fehlern liefern Ward & Clarksen. Sie unterscheiden in der Gruppe „gerätebezogene Fehler“<sup>52</sup> die Fehler, welche dem Hersteller zuzusprechen sind, von den menschlichen Fehlern [406], jedoch wird hierbei wiederum ein Versagen der Mensch-Maschine-Interaktion nicht explizit betrachtet. Derartige Fehler werden aus ihrem Blickwinkel entweder nur dem Design oder nur dem Menschen zugeordnet (Abb. 72).

---

<sup>52</sup> Fehler, bei denen medizinische Geräte beteiligt sind



**Abb. 72: Klassifizierung gerätebezogener Fehler**

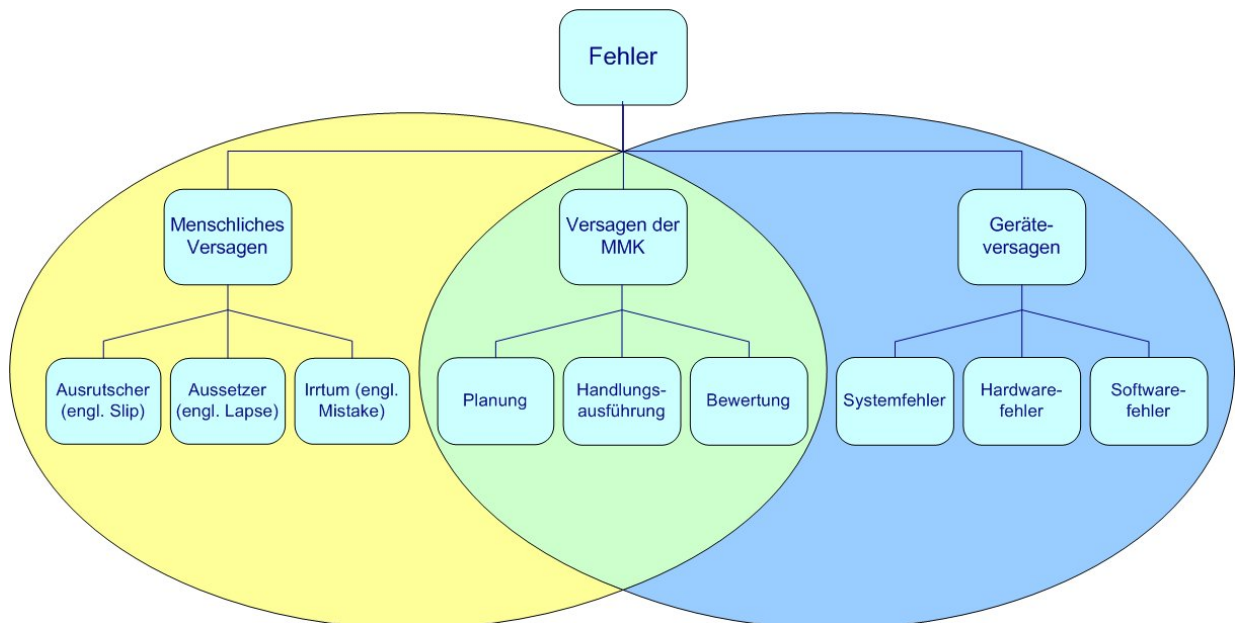
Um eine für diese Arbeit eindeutige Klassifizierung zu ermöglichen, wird im Folgenden eine Methode vorgestellt, welche die Fehlerklassifizierung der Mensch-Maschine-Interaktion in die medizinische Fehlerklassifizierung einbindet.

## 5.2 Erstellung eines Fehlerklassifizierungssystems unter Berücksichtigung der Mensch-Maschine-Schnittstelle

Die Kriterien für medizinische Fehler sind vielfältig. Sie unterscheiden sich u.a in ihrer Ursache, Art, Dauer und Auswirkung. Um die gesamte Vielfalt zu erfassen und ihre Bedeutung für den jeweiligen Schaden beurteilen zu können, ist es notwendig, jeden Fehler für sich genau zu analysieren und zu klassifizieren. Es gibt in der Literatur nur wenige Klassifizierungssysteme, die Fehler in technischen Systemen unter Berücksichtigung eines Versagens in der Mensch-Maschine-Interaktion klassifizieren [366]. Meist werden lediglich rein menschliche Fehler oder Maschinenfehler behandelt, weniger Mensch-Maschine-Fehler und so gut wie gar nicht das komplexe Gefüge. Erste Ansätze sind z.B. beim BfArM zu finden, jedoch werden hier menschliche Fehler ausgegrenzt, da sich das BfArM nur mit Fehlern in Zusammenhang mit den Geräten beschäftigt. Eine weitere Klassifizierung ist bei Montag zu finden [273]. Sie unterteilt medizinische Fehler im Zusammenhang mit der Gerätebenutzung in technische Einzelfehler, technische systematische Fehler, Anwendungsfehler, Anwenderfehler und Fehler, welche auf dem derzeitigen Stand der Technik beruhen. Auch diese Unterteilung ist eher oberflächlich. Daher wird im Folgenden eine Klassifizierung aufgestellt, die Fehler allgemein nach der Ursache einordnet (Abb. 73). Sie stützt sich auf die bereits in der Literatur etablierten Klassifizierungssysteme in den einzelnen Bereichen. Sie soll als eine erste Klassifizierung dienen. Für eine genauere Klassifizierung sei dann auf die entsprechenden tiefergehenden Klassifizierungssysteme der einzel-



nen Sparten verwiesen. Das angewandte Schema unterteilt in menschliches Versagen und Geräteversagen. Treffen diese beiden Komponenten zusammen, so wird von einem Versagen der Mensch-Maschine-Kommunikation gesprochen.



**Abb. 73: Fehlerklassifizierungsschema nach der Ursache, unter Berücksichtigung der Mensch-Maschine-Interaktion**

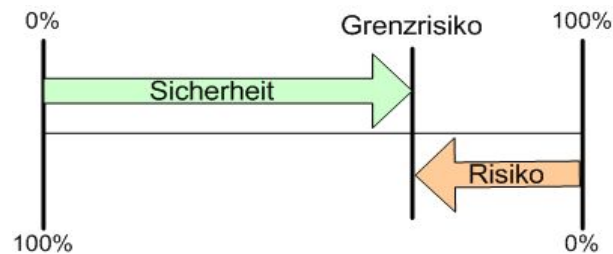
### 5.3 Sicherheit und sicherheitsbezogene Begriffe

*Keep people away from equipment  
until it becomes 'safe'* N. Storey

Im Gegensatz zu vielen anderen industriellen Bereichen muss bei der Betrachtung von Gebrauchstauglichkeitsaspekten in der Medizin besonderes Augenmerk auf das Thema „Sicherheit“ gelegt werden. Die besondere Sensibilität für das Thema Sicherheit in der Medizin beweist sich insbesondere in den in Kapitel 2.2 aufgeführten erschreckenden Statistiken. Wie bereits schon im IOM Report geäußert, hinkt der medizinische Sektor in Bezug auf die Gewährleistung von Sicherheit zehn oder mehr Jahre vergleichbaren Bereichen der Industrie hinterher [220], wo bemerkenswerte Fortschritte auf den Gebieten der Fehlerreduktion und der Sicherheitsgewährleistung gemacht wurden [366].

Sicherheit ist „die Freiheit von Risiko“ [343], ein „gefahrloser Zustand“ bzw. „ein gefahrloses Verhalten“ [309] und „die Freiheit von unfallbedingten Verletzungen“ [220]. Weitere Definitionen variieren in der Literatur. So sind beispielsweise Definitionen als ein in Erfahrung gegründetes und sich bestätigendes „Gefühl“, von

gewissen Gefahren nicht vorrangig getroffen zu werden [307], als eine „Sachlage“, bei der das Risiko nicht größer als das Grenzkrisiko ist (Abb. 74) [86] oder auch als eine „Bedingung“ für die Inbetriebnahme einer Anlage [86] zu finden.



**Abb. 74: Zusammenhang zwischen Sicherheit und Risiko**

Angesichts dieser unterschiedlichen Formulierungen ist Peters & Meyna [309] zuzustimmen, dass es keine von allen Institutionen und Fachleuten anerkannte Definition für den fundamentalen Begriff „Sicherheit“ gibt. Grundlegende Tatsache ist, dass Sicherheit nie absolut erreicht wird und somit auch nicht bewiesen werden kann, lediglich kann Unsicherheit definitiv festgestellt werden. Man spricht hierbei vom Paradoxon der „absoluten Sicherheit“ [341]. Eine weitaus praktikablere Definition liefern die für den medizinischen Bereich relevanten Normen DIN EN ISO 14971 und DIN EN IEC 60601-1-6 als die „Freiheit von nicht akzeptablen Risiken“ [97. 108]. Durch die Ergänzung von „nicht akzeptabel“ wird die Unwahrscheinlichkeit des Erreichens entschärft und die Definition praktikabler. Ein medizinisches Gerät kann somit zwar menschliches Leben gefährden, jedoch aufgrund des erhöhten Nutzens für das menschliche Leben ein akzeptables Risiko darstellen. Hierbei ist immer abzuwägen, was vertretbare Risiken sind.

Eine weitere, der DIN EN ISO 14971 ähnliche Definition, jedoch ausführlicher für den Aspekt der Sicherheit, liefert die DIN EN 61508. Demnach ist Sicherheit die „Freiheit von unverträglichen Risiken der physischen Verletzung oder Schädigung der Gesundheit von Menschen, entweder direkt oder indirekt als ein Ergebnis von Schäden an Gütern oder Umwelt“ [100]. Zusätzlich wird in der Norm der Begriff „Funktionale Sicherheit“ als Teil der Gesamtsicherheit eingeführt, der davon abhängig ist, dass ein System oder Betriebsmittel korrekte Antworten auf einen Eingangszustand liefert. Hinderungsgründe für das Erreichen funktionaler Sicherheit sind nach der DIN EN 61508 neben Hardwareausfällen und Softwarefehlern auch menschliches Versagen und demnach auch Fehler in der Mensch-Maschine-Interaktion [100]. Demzufolge kann Sicherheit in drei verschiedene Kategorien eingeteilt werden. Man unterscheidet die „primäre Sicherheit“, welche

sich auf die Systeme selbst bezieht. Das Risiko geht hierbei von Gefahren aus, welche durch das System direkt verursacht werden, wie z.B. Stromschläge, Verbrennungen und Brände. Die „funktionale Sicherheit“ beinhaltet Gerätekomponenten, die vom System kontrolliert werden. Sie hängt also von der korrekten Hard- und Softwarefunktion ab. Die dritte Kategorie ist die „indirekte Sicherheit“. Hierzu gehören beispielsweise falsche Information, Bearbeitungsdauer und auch die für diese Arbeit relevanten Mängel der Mensch-Maschine-Schnittstelle [380]. Weiterhin ist hervorzuheben, dass Sicherheit nicht die Eigenschaft einer speziellen Handlung oder Aufgabe ist, sondern die Zusammenhänge eines Arbeitssystems umfasst.

In Abb. 75 wird der Zusammenhang der wichtigsten sicherheitsbezogenen Begriffe in Anlehnung an Zimolong dargestellt.

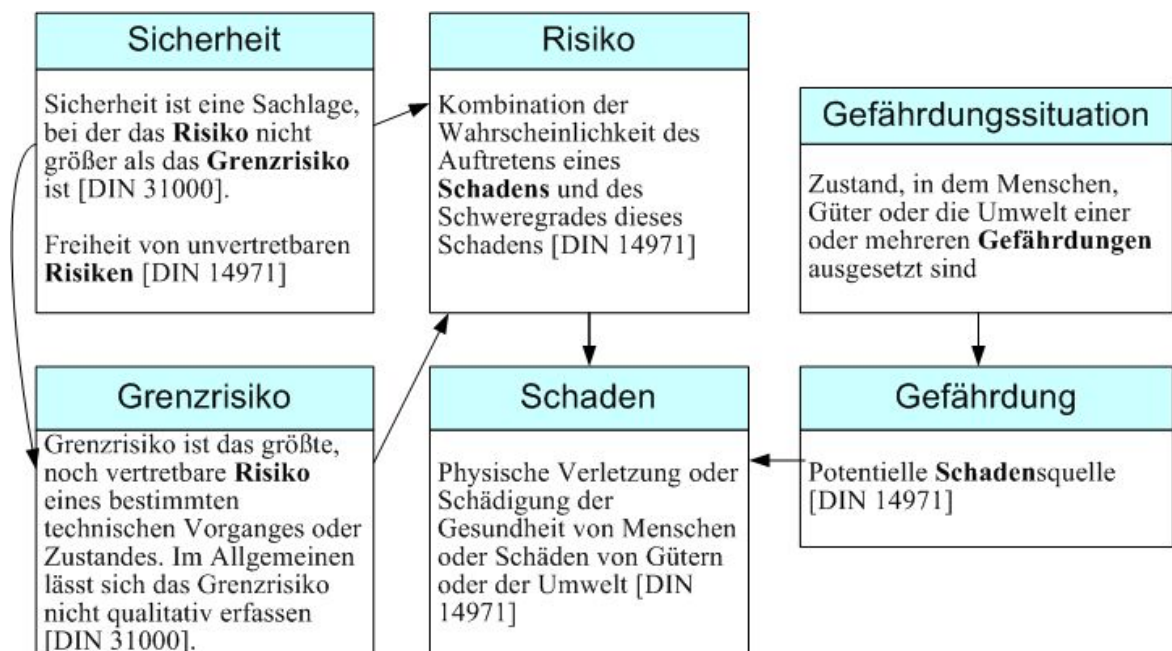


Abb. 75: Zusammenhang sicherheitsbezogener Begriffe in Anlehnung an Zimolong [432]

Neben den bereits beschriebenen qualitativen Definitionen von Sicherheit lassen sich auch quantitative Definitionen u.a. bei Peters & Meyna finden [309].

Kriterien für Sicherheit sind nach Storey: Zuverlässigkeit, Verfügbarkeit, Störungssicherheit, Systemintegrität, Datenintegrität, Systemwiederherstellung und Wartungsfreundlichkeit [380]. Diese Attribute sind jedoch lediglich auf Maschinen ausgelegt. Wie in Kap. 1.2.2.1 beschrieben, bildet das medizinische Arbeitssystem ein untereinander abhängiges Gesamtgefüge aus Mensch, Maschine und Aufgabe in einer gewissen Umwelt. Die Sicherheit kann hierbei durch jede einzelne Komponente gefährdet werden. Somit sollte Sicherheit nicht allein auf die

Maschine ausgerichtet werden, sondern auf das gesamte Mensch-Maschine-System und auf die darin stattfindende Interaktion. In dieser Mensch-Maschine-Interaktion kann Sicherheit im Sinne der DIN EN 9241-110 mit Hilfe der dort beschriebenen Anforderungen an die Dialoggestaltung [114] und durch ein sicheres Verhalten der Komponente Mensch erreicht werden. Demnach wird Sicherheit durch die Summe aus Systemsicherheit, Interaktionssicherheit und sicherem menschlichem Verhalten erreicht. Zusätzlich sollten bei der Interaktion in einem Mensch-Maschine-System beide Komponenten in der Lage sein, Fehler der jeweils anderen Komponente zu erkennen, abzufangen und zu korrigieren. Weitere zu beachtende ergonomische Aspekte bei der Sicherheit von Maschinen wie z.B. Körpermaße, -haltung, -bewegung und -kraft der bedienenden Person fallen in den Bereich der Anthropometrie (vgl. Kap. 2.1.2) und werden z.B. in der DIN EN 614-1 spezifiziert [99].

Die für den medizinischen Kontext spezielle Form der Sicherheit ist die Patientensicherheit. Sie ist die Vermeidung unerwünschter Ereignisse bei der ansonsten dem Wohl und der Gesundheit des Patienten dienenden medizinischen Behandlung. Nicht jedes unerwünschte Ereignis ist allerdings vermeidbar. Als unerwünscht im Sinne der Patientensicherheit kann ein Ereignis angesehen werden, dem ein Fehler vorausgegangen ist. Die Definition der Patientensicherheit nach der „National Patient Safety Foundation“ (NPSF) als die „Vermeidung, Prävention und Verbesserung der negativen Ergebnisse oder Verletzungen, die sich aus dem medizinischen Prozess ergeben“ [384] ist demnach auf vermeidbare Ereignisse zu beziehen.

Unter dem Begriff „Patientensicherheit“ hat sich in Deutschland eine wissenschaftliche Bewegung entwickelt, die sich mit diesem Thema beschäftigt und das medizinische Arbeitssystem im Hinblick auf Sicherheit erforscht. Hierbei ist das „Aktionsbündnis Patientensicherheit e.V.“<sup>53</sup> hervorzuheben. Diese Vereinigung aus Vertretern aller Gesundheitsberufe und -institutionen, Patientenorganisationen und Interessensvertretern steht seit April 2005 im Dienst einer sicheren Gesundheitsversorgung in Deutschland mit Strategien zur Vermeidung unerwünschter Ereignisse.

---

<sup>53</sup> [www.aktionsbuendnis-patientensicherheit.de](http://www.aktionsbuendnis-patientensicherheit.de) (08.12.2009)

### 5.3.1 Sicherheitskritische Systeme

Systeme, welche Sicherheitsfunktionen ausführen, werden sicherheitsbezogene [101] oder auch sicherheitskritische Systeme genannt (engl. *safety critical systems*) [35, 100, 380]. Börcsök versteht darunter ein System, das eine oder mehrere Funktionen zur Umsetzung der Anforderungen nach funktionaler Sicherheit besitzt, und dessen Versagen gefährliche Auswirkungen zur Folge hätte [35]. Nach der Auffassung vieler Autoren gilt ein System als sicherheitskritisch, wenn dessen Ausfall zu Verletzungen, Tod oder schwerwiegenden Umweltschäden führen kann [197, 300, 373]. Diese Definitionen beinhalten eine große Bandbreite von Equipment, von einem einfachen Mikrocontroller bis hin zu einem komplexen Atomkraftwerk. Derartige Systeme sind beispielsweise im Verkehrswesen, in der Raumfahrt, in der Energietechnik und im militärischen Bereich zu finden. Aufgrund der sicherheitsbedingten Relevanz zählen auch die meisten medizinischen Geräte zu den sicherheitskritischen Systemen. In diesen Systemen hinterlassen Fehler oft schwere Schäden bis hin zu Todesfolgen. Ausfälle in solchen sicherheitskritischen Systemen können nach Sommerville drei Ursachen haben: 1. Ausfälle der Hardware durch Entwurfsfehler oder Komponenten, die sich dem Ende ihrer Lebensdauer nähern, 2. Ausfall der Software aufgrund von Spezifikations-, Entwurfs- oder Implementierungsfehlern und 3. Bedienfehler als vorrangigste Ursache [373].

Gerade die letzte Problematik ist nie vollkommen zu beherrschen. Daher ist auch bei größtmöglicher Sicherheitsbeachtung und bestem Risikomanagement Storey beizupflichten, dass kein System absolut sicher sein kann [380]. In jedem System wird es weiterhin Restrisiken geben. Ziel sollte es sein, ein System zu entwerfen, in welchem die Restrisiken weitestgehend eingeschränkt, überschaubar und für die Arbeitsaufgaben angemessen und ausreichend sicher sind. Nach der DIN EN ISO 14971 ergeben sich somit Anforderungen an die Gestaltung des Systems im Sinne von: Beseitigung einer Gefährdung, Verringerung der Wahrscheinlichkeit des Schadenseintritts oder Verringerung der Schwere des Schadens [108]. Um dieses Ziel zu erreichen, muss auch die Gebrauchstauglichkeit betrachtet werden, insbesondere im Hinblick auf die von Sommerville vertretene Auffassung, dass fehlende Gebrauchstauglichkeit die häufigste Ursache für Ausfälle sicherheitskritischer Systeme ist.

Durch immer komplexer werdende Geräte mit ständig zunehmenden Bauteilen, Baugruppen und Funktionalitäten wird die Sicherheitsbetrachtung laufend schwie-

riger. Grundsätzlich sind Funktionsanzahl und Sicherheit zunächst zwei entgegengesetzte Eigenschaften von technischen Geräten [309]. Ähnlich unterliegt der Mensch bei gleichzeitiger Beanspruchung durch eine große Anzahl von Aufgaben in einem Mensch-Maschine-System einem statistisch erfassbaren, steigenden Fehlverhalten [309]. Ungefähr bei 5000 hochautomatisierten Handlungen macht der Mensch einen Fehler [355]. Diese statistische Fehlerrate wird durch Faktoren wie z.B. Stress zusätzlich erhöht. Hinzu kommen Fehler, welche durch Gebrauchstauglichkeitsprobleme verursacht oder begünstigt werden. Je komplexer die Systeme sind und je mehr Handlungsschritte durchgeführt werden, desto wichtiger wird die Sicherheit. Der Meinung vieler Hersteller, dass Sicherheit ein Feature oder eine Marketingstrategie sei, steht die Forderung nach Sicherheit in der Medizin als absolute Notwendigkeit entgegen. Daher verpflichtet auch das MPG in §7 den Hersteller eines Medizinproduktes, dafür zu sorgen, dass weder die Sicherheit der Patienten noch die Gesundheit der Anwender oder Dritter gefährdet wird, wenn das Produkt unter der vorhergesehenen Bedienung und Zweckbestimmung eingesetzt wird. Dabei müssen etwaige Risiken verglichen werden und angesichts der nützlichen Wirkung für den Patienten vertretbar bzw. mit einem hohen Maß des Schutzes von Gesundheit und Sicherheit vereinbar sein [360].

#### **5.4 Methoden zur Analyse der Sicherheit**

Die meisten Menschen können Risiko und Sicherheit nur schwer erfassen. Dies liegt zumeist daran, dass sie die beeinflussenden Faktoren nicht vollständig oder gar nicht ergründen können. Als erschwerender Faktor kommt hinzu, dass der Mensch in der Regel emotional geprägt und oft unlogisch in der Interpretation von Ereignissen ist [380]. Viele Menschen haben Angst vor dem Fliegen, obwohl die Fahrt zum Flughafen statistisch gesehen gefährlicher ist [83]. In vielen Bereichen und komplexen Systemen sind durch zahlreiche quantitative und qualitative Analysetechniken und Methoden die Abschätzung der Sicherheit und die Bestimmung des Risikos möglich. Quantitative Methoden der Risikoanalyse ermitteln durch statistische Untersuchungen Häufigkeit und Auswirkungen bestimmter Ereignisse. Sie sind besonders geeignet, wenn, wie leider so oft in der Medizin, das Schadensausmaß hoch ist. Qualitative Verfahren dienen in den meisten Fällen als Ergänzung der quantitativen Verfahren. Eine gute Zusammenfassung der Metho-

den lässt sich beispielsweise bei Peters [309] oder im Anhang E der DIN EN ISO 14971 [108] finden.

Das medizinische Arbeitssystem stellt hohe Anforderungen an die Sicherheit seiner Geräte. Sicherheit heißt jedoch nicht im absoluten Sinne, dass das Risiko gleich Null ist. Vielmehr ist ein Produkt als sicher anzusehen, wenn das Risiko gegenüber dem erwarteten Nutzen angemessen ist [396]. Somit sind Definitionen, die Sicherheit als „die Freiheit von Risiko“ [209] betrachten, nur so zu deuten, dass die in der Technik nur theoretisch bestehende vollkommene Sicherheit gemeint ist. Ansonsten wäre diese Definition falsch. Sicherheit steht im direkten Zusammenhang mit dem zu erbringenden Aufwand. Je höher die Sicherheit, desto mehr Aufwand muss erbracht werden.

Risiko und Sicherheit werden im Rahmen des Risikomanagements erfasst. Als Schlüsselkomponente des Qualitätsmanagements umfasst das Risikomanagement im weitesten Sinne alle Tätigkeiten, Prozesse oder Methoden, die zur Risikoerfassung, -evaluation, -bewertung, -reduzierung und -kontrolle angewendet werden [212]. Nach Erfassung und Bewertung der Risiken ist der Schwerpunkt des Risikomanagements die Risikobeherrschung oder -minderung [212]. Um Risiken zu bewerten, müssen diese zuerst analysiert werden. Hierzu muss der Begriff des Risikos zunächst erfasst werden. Oft wird Risiko durch negative Ereignisse beschrieben. Das Bundesinstitut für Risikobewertung<sup>54</sup> definiert Risiko als „die Wahrscheinlichkeit, mit der ein möglicher Schaden eintritt“. Bei vielen Definitionen wird neben der für ein Risiko wichtigen „Wahrscheinlichkeit des Auftretens“ zusätzlich auch noch das Schadensausmaß mit einbezogen [108, 212, 234, 380]. Daher wird generell Risiko als eine Verkettung verschiedener Umstände angesehen. Abb. 76 verdeutlicht dies. Risiko wird als Summe von „Gefahrenstärke“, „Eintrittswahrscheinlichkeit“, „Gefahrenaussetzung“ und „Auftrittswahrscheinlichkeit eines Unfalls aufgrund einer Gefahr“ dargestellt.

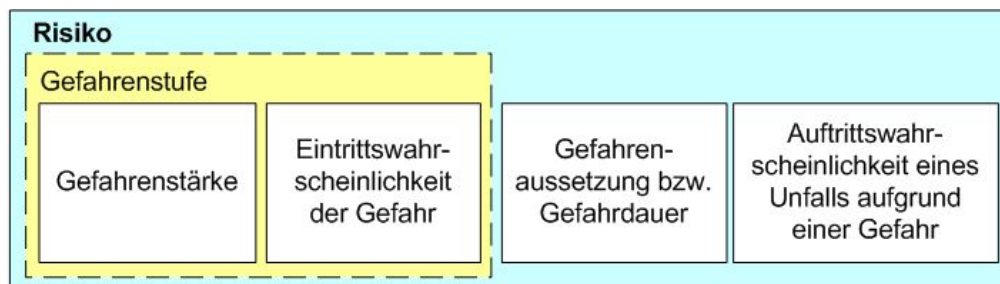


Abb. 76: Risikoelemente [234]

<sup>54</sup> <http://www.bfr.bund.de/cd/9110> (09.10.2008)

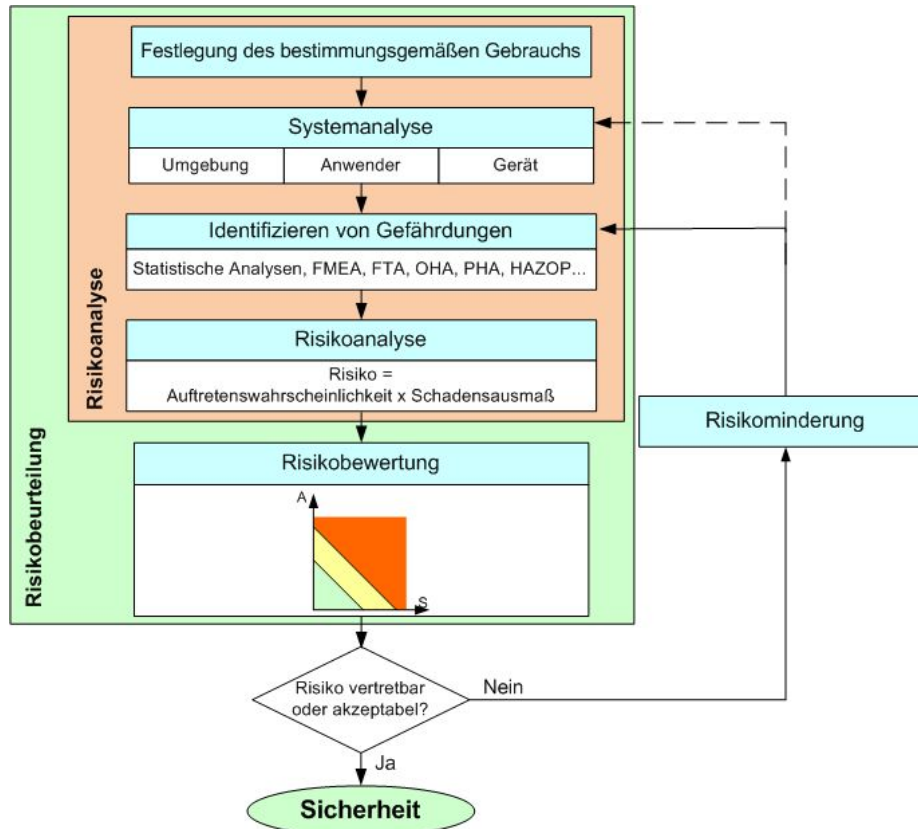
Gerade die Verkettung verschiedener Umstände stellt ein großes Problem bei der Risikoabschätzung dar. Bei der Bedienung eines medizinischen Gerätes wird das Risiko beispielsweise durch die Wahrscheinlichkeit, dass das Gerät aktiv benutzt wird, die Wahrscheinlichkeit, dass der Benutzer falsch reagiert, die Wahrscheinlichkeit, dass die Situation nicht bemerkt wird und durch die Schwere der Konsequenz beeinflusst. Meistens handelt es sich bei der Risikoabschätzung um sehr komplexe und schwer zu modellierende Zusammenhänge [234]. Die Risiko- und Gefahrenanalysen, welche auf der mathematischen Theorie der Wahrscheinlichkeit basieren, sind hierbei hilfreiche Methoden [393].

#### **5.4.1 Risikoanalyse**

Die Risikoanalyse, auch Gefahrenanalyse genannt, dient zur Identifikation von Gefahren. Zusätzlich beschäftigt sie sich mit der Identifikation von Umweltbedingungen und Gefahrenauswirkungen [234]. Sie bildet zusammen mit der später beschriebenen Risikobewertung die Risikobeurteilung. Für medizinische Geräte werden die Vorgänge der Risikoanalyse innerhalb des Risikomanagementprozesses für Medizinprodukte durch die DIN EN ISO 14971 geregelt (vgl. Kap. 2.3) [108].

Die Risikoanalyse nach der DIN EN ISO 14971 ist in drei Phasen eingeteilt. Sie betrachtet neben dem bestimmungsgemäßen Gebrauch eines Erzeugnisses, Verfahrens oder einer Dienstleistung auch den „vernünftigerweise vorhersehbaren Missbrauch“ [108]. Der Leitfaden fordert eine Risikominderung durch Anwendung des iterativen Verfahrens der Risikobeurteilung (Abb. 77).





**Abb. 77: Iteratives Verfahren von Risikobeurteilung und Risikominderung [97, 108]**

Als Erstes muss der Hersteller den bestimmungsgemäßen Gebrauch (vgl. Kap. 2.3) und den vorhersehbaren Missbrauch definieren. Es müssen alle Merkmale beschrieben werden, welche die Sicherheit des Medizinproduktes beeinflussen können und somit risikoprägend sind. Um diese zu identifizieren, gibt die Norm einige Fragestellungen vor, die den Anwender einer Risikoanalyse bei der Identifizierung von Risiken unterstützen sollen. Beispiele sind:

- Welches ist die Zweckbestimmung, und wie soll das Medizinprodukt angewendet werden?
- Ist es vorgesehen, dass das Medizinprodukt mit dem Patienten oder anderen Personen in Berührung kommt?
- Liefert das Medizinprodukt interpretierende Aussagen?
- Können die Gestaltungsmerkmale der Schnittstelle zum Anwender zu den Fehlern bei der Anwendung beitragen?

Ein weiterer Fragenkatalog, der bei einer Risikoanalyse zur Identifizierung von Gefährdungen, besonders unter dem Gesichtspunkt der sicherheitskritischen Bedienung, hilfreich sein kann, wurde von der Kommission Arbeitsschutz und

Normung in dem „Ergonomie-Leitfaden für die Gestaltung von Medizinprodukten“ herausgegeben [186].

Nach Bestimmung dieser Merkmale werden die bekannten und vorhersehbaren Gefährdungen beschrieben. Hierbei soll ebenfalls auf Gefährdungen eingegangen werden, welche mit dem Medizinprodukt bei Fehlbedienung in Zusammenhang stehen. In der letzten Phase müssen die Risiken für jede Gefährdung eingeschätzt werden. Die Risikoeinschätzung erfolgt hierbei durch eine Kombination der Komponenten Wahrscheinlichkeit und Schweregrad des Schadens. Die Verwendung der Verfahren schränkt die Norm hierbei nicht ein, lediglich sollten die Komponenten getrennt analysiert werden. Als Hilfe für das Bewertungsniveau schlägt die Norm ein Risiko-Diagramm vor, welches das Risiko in drei Bereiche einstuft, um dessen Vertretbarkeit zu bestimmen (►Kap. 5.4.3). In einigen Fällen werden Risiken so niedrig sein, dass sie beispielsweise im Vergleich mit dem Nutzen vernachlässigbar sind. Bei nicht vertretbarem Risiko muss ein Risikoreduzierungsprozess angestoßen und nach Änderung des Risikoprozesses von Neuem begonnen werden.

Generell ist die DIN EN ISO 14971 in den Bereichen der Risikoeinschätzung und -bewertung wenig konkret. Exaktere Angaben bezüglich dieses Prozesses lassen sich in der Norm zur „Sicherheit von Maschinen – Leitsätze zur Risikobeurteilung“ DIN EN 1050 [94] in Abschnitt 7 „Risikoeinschätzung“ und Abschnitt 8 „Risikobewertung“ finden. In Abschnitt 7.3.4 werden explizit menschliche Faktoren als risikobeeinflussend genannt. Hierbei wird auch die Wechselwirkung zwischen Personen und Maschinen, also die Mensch-Maschine-Kommunikation, mit einbezogen. Zur Einschätzung der Fähigkeiten von gefährdeten Personen müssen Gesichtspunkte wie ergonomische Gestaltung, Risikobewusstsein, Vertrauen auf absichtsgemäße Ausführung der geforderten Aufgabe, Anreize zur Abweichung von den notwendigen oder vorgeschriebenen Arbeitsverfahren etc. berücksichtigt werden. Die Norm betont ausdrücklich, dass Ausbildung, Erfahrungen und Fähigkeiten das Risiko beeinflussen können, aber nicht als Argument verwendet werden dürfen, um als Ersatz für eine Risikominderung zu dienen, welche durch eine konstruktive oder technische Schutzmaßnahme hätte erreicht werden können [94].

## **Methoden der Risikoanalyse**

Neben den bereits beschriebenen quantitativen Analyseverfahren sind qualitative Verfahren insbesondere zur Identifizierung bestimmter Gefahren und Handlungen üblich. Die Gefahrenanalyse (engl. Hazard Analysis) identifiziert die potentiellen Möglichkeiten, welche Schaden anrichten können, und weist sie einer Gefahrenstufe zu. Erkannte Probleme werden auf ihre Wichtigkeit hin untersucht. Für eine Gefahrenanalyse steht ein vielfältiges Methodenrepertoire zur Verfügung, welches in der Regel darauf abzielt, Abläufe, die zu einem Unfall führen könnten, zu identifizieren. Diese Folge von Ereignissen wird oft als Unfallsequenz oder Unfallszenario bezeichnet [380]. Die Vielzahl der Techniken kann in Vorwärts- und Rückwärtssuche oder Top-Down- und Bottom-Up-Suche eingeteilt werden [35]. Bei den Techniken wird die Betrachtung nicht auf einzelne Komponenten des Systems gerichtet, sondern das System wird als Ganzes betrachtet. Es wird versucht, im Gesamtüberblick Gefahren und ihre Ursachen zu erkennen, um diese zu reduzieren oder auszuschließen. Börczek empfiehlt die Verwendung mehrerer Techniken, denn die dadurch erzielte Berücksichtigung verschiedener Sichtweisen der Gefahr kann zu neuen Erkenntnissen führen [35]. Einige ausgewählte Verfahren werden im Folgenden beschrieben:

### **FMEA (Failure Mode and Effect Analysis)**

Die in Deutschland auch „Ausfallart- und Fehleranalyse“ [309], oder „Fehlermöglichkeits- und Einflussanalyse“ genannte Methode FMEA ist eine Qualitätssicherungsmethode. Durch eine systematische und vorausschauende Analyse sollen mögliche Fehler und potentielle Schwachstellen eines Produktes oder Prozesses erkannt und deren Auswirkungen bewertet werden. Es sollen präventive Maßnahmen entwickelt werden, um das Auftreten von möglichen Fehlern bereits im Vorfeld der Design- und Entwicklungsphasen zu vermeiden oder zu reduzieren. Die erstmalig von den US Streitkräften angewandte Methodik hat sich heute in vielen Bereichen der Industrie etabliert und wird häufig im Automobilbereich eingesetzt [79, 234]. Mit entsprechenden Anpassungen kann sie problemlos in anderen Industriezweigen angewendet werden. Auch lassen sich Methoden mit der Bezeichnung „Healthcare Failure Mode and Effects Analysis“ (HCFMEA) [380] finden. Diese suggerieren zwar eine direkte Ausrichtung auf den Gesundheitsmarkt, unterscheiden sich jedoch im Wesentlichen nicht von der Ursprungsmethode.

Die FMEA wird meist von Experten in einem interdisziplinären Team in folgender Weise durchgeführt [79, 309]:

1. Systemeingrenzung und Strukturierung,
2. Definition von Funktionen und Strukturelementen,
3. Analyse potentieller Fehlerursachen, Fehlerarten und Fehlerfolgen,
4. Risikobewertung durch Ermittlung der Auftretenswahrscheinlichkeit (A), Bedeutung (B) und Wahrscheinlichkeit der Entdeckung (E) vor Auslieferung an den Kunden, mit jeweiligen Bewertungsstufen von „1“ = gering bis „10“ = hoch,
5. Entwicklung von Maßnahmen und Lösungsvorschlägen.

Durch Multiplikation wird eine Risikoprioritätszahl (RPZ) zur Festsetzung der Priorität eines Problems ermittelt ( $RPZ = B \times A \times E$ ). Je höher die RPZ, desto höher die Priorität. Grundsätzlich ist es Unternehmen freigestellt, ab welcher Grenze sie Verbesserungsmaßnahmen anstreben, eine gängige Schranke ist jedoch ein RPZ-Wert von 125 [353].

### **Human FMEA**

Bei der FMEA wird der Faktor Mensch in der Sicherheitsbetrachtung weitgehend ignoriert. Wegen dieses Defizits ist die Human FMEA als eine besondere Form der FMEA anzusehen. Sie bezieht den Menschen mit in die Risikobetrachtung ein und untersucht Fehler, welche durch menschliches Verhalten hervorgerufen werden [309]. In der Literatur lassen sich hierzu zwei verschiedene Ansätze finden, von Helbling & Weinbeck, sowie von Algedri & Frieling. Der von Helbling & Weinbeck erarbeitete Ansatz ergänzt die klassische FMEA um eine zusätzliche Ebene, d.h. es werden neben der Fehlerfolge, dem Fehler und der Fehlerursache mögliche menschliche Einflussfaktoren, welche für die Handlungsfehler verantwortlich sind, aufgelistet. Das Risiko wird mittels der klassischen Risikoprioritätszahl bewertet. Ergänzend werden hierbei die Arbeitsprozesszuteilung, Handlungsfehler, Handlungsphase, Einflussfaktoren, Beeinflussbarkeit und Fehlerform betrachtet [177]. Der zweite Ansatz gliedert sich in drei aufeinander aufbauende Phasen [1]. In der ersten Phase, der Produktfehleranalyse, werden die relevanten Produktfehler sowie fehlerträchtige Prozesse und Handlungen erfasst, klassifiziert, visualisiert und bewertet. In der zweiten Phase, der Handlungsfehleranalyse, werden Ursachen bzw. fehlerauslösende Bedingungen und Handlungen ermittelt, um dann in der letzten Phase Gestaltungsanforderungen abzuleiten.

Hierzu werden für jede Phase spezifische Werkzeuge zur Erfassung, Auswertung bzw. Verarbeitung der jeweiligen Phase beschrieben. Auch bei dieser speziellen Ausprägung der FMEA wird eine Kooperationsbereitschaft von Angestellten und Führungskräften vorausgesetzt.

### FTA (Fault Tree Analysis)

Die „Fault Tree Analysis“, oder zu deutsch „Fehlerbaumanalyse“ ist eine Methode, um die Wahrscheinlichkeiten von Ausfällen bestimmen zu können. In Deutschland regelt die Norm DIN 25424 den Ablauf einer Fehlerbaumanalyse [85]. Sie stellt eine grafische Methode zur Ermittlung von logischen Verknüpfungen zwischen Komponenten- oder Teilsystemausfällen dar [35], startet mit einem Event, welches direkt mit einer Gefahr (hazard) verbunden ist ('top event') und erarbeitet schließlich nach dem Top-Down-Prinzip die möglichen Ursachen. Dazwischenliegende Events werden mit logischen Operanden (AND oder OR) verknüpft, welche in der Norm DIN EN 61025 festgelegt sind [98]. Ein einfaches Beispiel eines Fehlerbaumes für die Schädigung durch eine Spritze ist in Abb. 78 dargestellt.

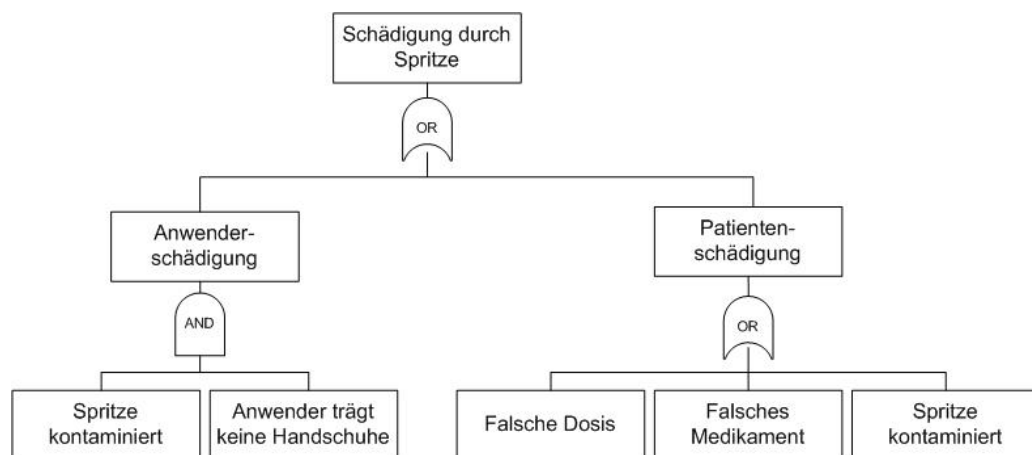


Abb. 78: Beispiel für einen einfachen Fault Tree

### OHA (Operating Hazard Analysis)

Die Operating Hazard Analysis (OHA) (zu deutsch: Bedienungsgefahrenanalyse) ermittelt Gefahren, die durch Bedienung, Wartung und Reparatur entstehen können [234, 309]. Im Fokus dieser Methode stehen Gefahren, welche aus Aufgaben und Aktivitäten bei der Verwendung des Gerätes hervorgehen, unter besonderer Berücksichtigung der Ergonomie [81]. Die OHA sollte so früh wie möglich in die Produktentwicklung eingebunden werden. Normalerweise wird die OHA durch eine Expertengruppe durchgeführt [335]. Alle Arbeitsaufgaben und menschlichen Handlungen, sowie Unterlassungshandlungen der interagierenden Personen,

sollten durch die Expertengruppe hinsichtlich Gefahren untersucht werden. Gefundene Gefahren werden bewertet und jeweilige Gegenmaßnahmen eingeleitet. Durch Fokussierung der menschlichen Tätigkeit wird mit Hilfe der OHA eine Minimierung der Fehler erreicht, welche durch den Menschen oder durch die Mensch-Maschine-Kommunikation verursacht wird. Sie bewirkt durch Verbesserungen in Design, Verfahren, Warnvorrichtungen, Handbüchern, Schulungen etc. eine höhere Effizienz und Sicherheit.

### **Preliminary Hazard Analysis (PHA)**

Die Preliminary Hazard Analysis (PHA) (zu deutsch: vorläufige Gefahrenanalyse) ist ein induktives Verfahren, um in allen Lebensphasen eines Systems die Gefährdungen, Gefährdungssituationen und Gefährdungsereignisse festzustellen [94, 234, 291, 309]. Die Methode besteht meist aus einem Brainstorming, in dem Experten in sicherheitskritischen Bereichen eine erste Einschätzung der Gefahren vornehmen und tabellarisch dokumentieren. Hierbei werden in der Regel jeder Gefahr mögliche Auswirkungen, Schweregrade und Auftretenshäufigkeiten zugeordnet sowie entsprechende Schritte zu ihrer Vermeidung vorgegeben (Tab. 26).

<b>Gefahr</b>	<b>Auswirkung (Unfall)</b>	<b>Schweregrad</b>	<b>Randbedingung</b>	<b>Gefahrenpotential</b>	<b>Gefahrenvermeidung</b>
Falsche Dosierung	Tod oder ernste Verletzung des Patienten	Kritisch	Kritischer Zustand des Patienten	Häufig	Unwahrscheinlich, dass die Gefahr vermieden werden kann, jedoch kann die Auswirkung verringert werden, wenn die Vitalparameter überwacht werden.
Falsches Medikament	Tod oder ernste Verletzung des Patienten	Kritisch	Kritischer Zustand des Patienten	Häufig	Unwahrscheinlich, dass die Gefahr vermieden werden kann, jedoch kann die Auswirkung verringert werden, wenn die Vitalparameter überwacht werden.

**Tab. 26: PHA am Beispiel einer Spritzenpumpe**

### ***Hazard and Operability Studies (HAZOP)***

Die Hazard and Operability Studies (HAZOP)-Methode wurde in den 60er Jahren von „Imperial Chemical Industries“ (ICI) entwickelt, um Gefahren in Chemieanlagen zu analysieren. Bis heute ist sie in diesem Bereich erhalten geblieben, wird aber auch in vielen weiteren Bereichen erfolgreich eingesetzt [291, 380]. Im deutschsprachigen Raum ist diese Methode unter dem Namen PAAG verbreitet. Die Buchstaben stehen hierbei für die jeweiligen Handlungsschritte:

1. **P**rognose eines möglichen Ereignisses
2. **A**uffinden der Ursachen
3. **A**bschätzen der Auswirkungen
4. **G**egenmaßnahmen

Unter der Leitung eines Moderators untersucht ein interdisziplinäres Expertenteam in verschiedenen Sitzungen durch systematisches Vorgehen ein System hinsichtlich eventuell auftretender Abweichungen und bestimmt entsprechende Gegenmaßnahmen. Um die Abweichungen in einem System darzustellen, werden einfache Leitwörter verwendet. Mit Hilfe dieser Leitwörter, wie z.B. mehr/weniger, kein, teilweise, anders als, werden Systemkomponenten verknüpft, um das System zu analysieren. Beispielsweise werden bei einer Verknüpfung mit dem Leitwort „kein“ und der Systemkomponente „Netzteil“, bezogen auf eine Spritzenpumpe, alle Abweichungen diskutiert, die auftreten können, wenn kein Stromanschluss am Netzteil ist. Die Ergebnisse werden tabellarisch dokumentiert. Beispiele sind in Tab. 27 dargestellt.

Leitwort	Abweichung	Mögliche Ursachen	Folgen	Notwendige Maßnahmen
KEIN	Netzanschluss	-Netzstecker versehentlich entfernt -Netzstecker für Transport entfernt -Netzstecker nicht gesteckt	Umschalten auf Akkubetrieb	1. Akkubetrieb anzeigen 2. Frühzeitige Information über Akkustatus 3. Notakku als eigenes System 4. Reserveakku
KEIN	Akkubetrieb	-Akku leer -Akku defekt -Akku nicht angeschlossen	Gerät schaltet sich aus, keine Förderung	1. frühzeitige Warnmeldung 2. Notakku als eigenes System 3. Reserveakku
MEHR	Medikament	-Fehlbedienung -Falscher Sprizentyp -Missbrauch -Hardwareproblem -Softwareproblem	Patient erhält eine zu hohe Dosis des Medikaments	1. Vitalparameter überwachen 2. Medikamentendatenbank anlegen 3. Defaultwerte definieren

Tab. 27: Exemplarische PAAG am Beispiel einer Spritzenpumpe

Problematisch bei den meisten Gefahrenanalysemethoden ist die Ausrichtung auf systematische und meist technische Fehler. Obwohl einige Methoden auf eine Designorientierung angepasst wurden, werden menschliche Fehler oder Fehler der Mensch-Maschine-Kommunikation meist nicht beachtet [146]. Bei genauerer Betrachtung der meisten Fehler ist jedoch festzustellen, dass begangene Fehler häufig auf den Menschen oder die Interaktion zwischen Menschen und der Maschine zurückzuführen sind, statt auf technische Fehler (vgl. Kap. 5.1.4). Über

80% der Arbeitsunfälle entstehen durch eine menschliche Aktion und nicht durch ein technisches Versagen [234]. Da der Mensch bei den meisten Arbeitsschritten entscheidend involviert ist, muss bei der Durchführung einer Gefahrenanalyse auch der menschliche Einfluss berücksichtigt werden. Dies stellt sich allerdings als schwierig heraus, da Einflussgrößen wie z.B. Müdigkeit, Stress und Motivation schwer zu bewerten und zu erfassen sind. Bei derartigen Bewertungen muss darauf geachtet werden, dass Menschen Individuen sind und dementsprechend unterschiedlich in bestimmten Situationen reagieren. Somit sollte auch bei einer Gefahrenanalyse nicht von einem stereotypen Menschen, sondern von unterschiedlichen Individuen mit entsprechenden Stärken und Schwächen ausgegangen werden. Daher können zwar Elemente dieser Methoden für die Einstufung der Gebrauchstauglichkeit medizinischer Geräte verwendet werden, für sich alleine reichen sie hierfür jedoch nicht aus.

#### 5.4.2 Risikobewertung

*Das Verhüten von Unfällen darf nicht als Vorschrift des Gesetzes aufgefasst werden, sondern als ein Gebot menschlicher Verpflichtung und wirtschaftlicher Vernunft*  
Werner von Siemens, 1880

Die Risikobewertung knüpft an die Risikoanalyse an. Sie beurteilt Gefahren mit Hilfe wissenschaftlicher Methoden unter Betrachtung der Umweltbedingungen, welche zu der Gefährdung führen, der Gefahrendauer, der Gefahrenaussetzung und der Gefährdungswahrscheinlichkeit [234]. Sie basiert auf der mathematischen Wahrscheinlichkeitstheorie.

Menschen sind ständig Gefahren ausgesetzt. Mit ihnen sind bestimmte Risiken verknüpft, deren Wichtigkeit abgeschätzt werden muss. Die Gefahr, von einem Blitz getroffen zu werden, ist ebenso präsent, wie bei einem Verkehrsunfall zu sterben, jedoch ist das entsprechende Risiko jeweils unterschiedlich. Die **Auftretenswahrscheinlichkeit** spielt hierbei eine besondere Rolle. Die Wahrscheinlichkeit, in seinem Leben von einem Blitz erschlagen zu werden,  $(1:20.000.000)^{55}$ , ist bedeutend geringer, als bei einem Verkehrsunfall (ca. 1:16.000) zu sterben. Ebenso wichtig ist das **Schadensausmaß**. Sind zwei Gefahren gleichwahrchein-

---

<sup>55</sup> Ausgehend von einem Land mit einer Population von 50 000 000 sterben pro Jahr 25 Menschen durch einen Blitz  $25/50\,000\,000 = 0,5 \cdot 10^{-7}$



lich, so stellt die Gefahr mit dem höheren Schadensausmaß ein höheres Risiko dar. Andererseits muss bei einem höheren Schadensausmaß nicht unbedingt das Risiko höher sein. Hierbei gibt es eine untrennbare Abhängigkeit zwischen Schadensausmaß und Auftretenswahrscheinlichkeit. Um ein Risiko bestimmen zu können, müssen somit die Gefahr und die Auftretenswahrscheinlichkeit in Beziehung gesetzt werden. Dies kann sowohl qualitativ als auch quantitativ erfolgen. Der quantitative Ansatz versucht, sämtliche Größen der Risikobewertung numerisch exakt zu berechnen. Der Vorteil hierbei liegt darin, dass die Ergebnisse der Risikoanalyse auf genauen Messgrößen basieren. Die vorhandenen Methoden zur Ermittlung der Messgrößen sind jedoch meist sehr aufwändig und in gewissen Bereichen nur schwer möglich. Für Ausfallwahrscheinlichkeiten mechanischer Teile oder Naturkatastrophen gibt es verlässliche Angaben. Wie groß jedoch die Wahrscheinlichkeit ist, dass ein Chirurg in einer Stresssituation, bedingt durch ein schlechtes Gerätedesign, einen Fehler macht, kann statistisch nur schwer erfasst werden.

Üblicherweise wird das Risiko „R“ als die Auswirkung oder Ausfallfolge eines unerwünschten Ereignisses pro Zeiteinheit definiert. Nach Peters sind jedoch als Bezugsgrößen auch andere Normierungen möglich [309], wie z.B. Ausfälle je Patient, je Stück, je Tag oder je Behandlung. Somit gilt folgende einfache Beziehung:

$R = H \cdot A$ , wobei:	H= Häufigkeit [Mittlere Anzahl der Ereignisse $E_i$ / Zeiteinheit] A= Ausfallfolgen [Auswirkungen / Ereignis]
Für das Risiko des i-ten Ereignisses folgt dann:	$R_i = h_i \cdot a_i$
Und schließlich bei Betrachtung n möglicher Ereignisse:	$R_n = \sum_{i=1}^n h_i \cdot a_i \leq R_{ak}$
( $R_{ak}$ = akzeptierbares Risiko)	

In vielen Fällen sind nicht genügend quantitative Daten vorhanden oder können nicht erhoben werden. In diesem Fall sollen qualitative Beschreibungen erstellt werden. Die DIN EN ISO 14971 stellt klar, dass eine gute qualitative Beschreibung gegenüber einer quantitativen Ungenauigkeit zu bevorzugen ist [108].

Die **qualitative Risikobewertung** ist einfacher und schneller durchzuführen als die quantitative. Bei ihr werden anstelle von genauen Zahlenwerten Abschätzungen in verschiedenen Skalen durchgeführt. Der Vorteil ist hierbei der relativ geringe Aufwand. Dennoch stellen die Schätzungen lediglich subjektive Werte dar. Die Klassifikation eines Risikos wird, wie bei der quantitativen Risikoanalyse, durch

den Bezug von Schwere zur Häufigkeit des Ereignisses vorgenommen, um auf Basis dieser Einschätzung dem Risiko eine Klasse zuzuweisen. Verschiedene Standards existieren bereits in der Luft- und Raumfahrt, der Atomindustrie sowie im militärischen Bereich [71, 102, 198, 267, 309, 380, 347]. Die Medizin, als ein weiterer Hochrisikobereich, hinkt beim Fehlermanagement jedoch noch weit hinterher [433]. Erst seit Kurzem ist die Thematik auch in diesem Bereich aktuell und gewinnt mehr Beachtung [108, 220].

Zur Einstufung des **Schweregrades** gibt es verschiedene Möglichkeiten. Die Standards DIN EN 61508 und DIN EN 61511-3 für sicherheitsbezogene Systeme klassifizieren Risiko zunächst auf der Basis der Konsequenzen [102, 103]. Hierbei werden vier Kategorien unterschieden: katastrophal, kritisch, marginal und unbedeutend. Genaue Bedeutungen der Kategorien werden vom Standard nicht vorgegeben und müssten für den entsprechenden Industriezweig bzw. für die jeweilige Situation angepasst werden [102]. In Tab. 28 sind die Kategorien mit möglichen Beispielen angegeben.

Kategorie	Mögliche Beispiele
katastrophal	vielfache Todesfälle
kritisch	ein einzelner Todesfall und / oder schwere Verletzungen oder schwere berufsbedingte Krankheit
marginal	eine einzelne schwere Verletzung oder leichte berufsbedingte Krankheit
unbedeutend	meistens eine einzelne leichte Verletzung oder leichte berufsbedingte Krankheit

Tab. 28: Einstufung des Schweregrades der Fehlerkonsequenzen [380]

Weitere Beispiele aus dem Militärbereich und der Luft- und Raumfahrt lassen sich in der Literatur finden [71, 267, 309, 347, 380]. Sie ähneln alle dem beschriebenen Prinzip der Risikoeinteilung in verschiedene Stufen. Für Medizinprodukte schlägt die DIN EN ISO 14971 ebenfalls eine Einteilung in Wahrscheinlichkeitsstufen nach Schlagworten (Deskriptoren) vor. Die Norm ist hierbei jedoch wenig konkret und überlässt dem Anwender die Freiheit einer individuellen Definition und Stufenwahl [108].

Auch bei der Bewertung der **Wahrscheinlichkeit** des Schadenseintritts ist sich die Literatur uneinig. Für medizinische Geräte lässt die DIN EN ISO 14971 ebenfalls dem Anwender die Freiheit, eigene Wahrscheinlichkeitsstufen zu definieren, wenn keine genauen quantitativen Daten vorliegen oder beschafft werden können [108]. Konkretere Beispiele für eine Einteilung nach der Wahrscheinlichkeit werden von der DIN EN 61508 und DIN EN 61511 [102, 103] durch eine Unterteilung

in sechs Klassen für Ereignishäufigkeit festgelegt, die je nach Bedarf frei definiert werden können (Tab. 29).

Unfallhäufigkeit	Ereignisse während des Betriebes, unter Einbeziehung aller Systembestandteile
Häufig	wahrscheinlich ständig auftretend
Wahrscheinlich	wahrscheinlich oft auftretend
Gelegentlich	wahrscheinlich einige Male auftretend
Gering	wahrscheinlich irgendwann auftretend
Unwahrscheinlich	unwahrscheinlich, aber in Ausnahmefällen auftretend
Nicht glaubhaft	in hohem Maße unwahrscheinlich, dass das Ereignis überhaupt eintritt

**Tab. 29: Wahrscheinlichkeitseinteilung nach der DIN EN 61508-5 [102]**

Die quantitativen Methoden können rechnerisch durch Schweregrad und Eintrittshäufigkeit das Risiko bestimmen. Qualitativen Methoden fehlen hierfür die entsprechenden objektiven Daten. Sie erfordern entsprechende Prozesse, um Schweregrad und Eintrittshäufigkeit einer Gefahr in angemessener Weise zu kombinieren. Die Festlegung der Klassen erfolgt durch verschiedene Methoden. Die bekanntesten sind die Risikomatrix und der Risikograph.

### **Risikomatrix**

Die Risikomatrix ist eine Methode, um ein Risiko zu bestimmen und in Klassen einzuteilen. Dieses Vorgehen wird in vielen Industrie- und Wirtschaftsbereichen angewandt. Hierbei ist es unerheblich, ob es sich um ein sicherheitsbedingtes, finanzielles oder gesundheitliches Risiko handelt. Bei dieser Methode werden Schweregrad und Eintrittswahrscheinlichkeit in angemessener Weise miteinander kombiniert, um Risikoklassen zu erhalten.

Die Zuweisung kann qualitativ mit Hilfe von Urteilsvermögen oder quantitativ mit Hilfe von Berechnungen oder Statistiken erfolgen. Die Festlegung der Klassen unterscheidet sich in den verschiedenen Standards und Methoden. Sie basieren aber auf dem gleichen Prinzip [101, 234, 267]. In der Literatur lassen sich viele Abstufungen finden, nicht selten in bis zu fünf Kategorien unterteilt. Tab. 30 verdeutlicht die klassische Kategorieneinteilung für Auswirkungen und Wahrscheinlichkeiten in der Risikobetrachtung. Im Folgenden wird die Risikomatrix am Beispiel der DIN EN 61508-4 [101] für sicherheitsbezogene elektrische Systeme vorgestellt. In der Norm werden vier Risikoklassen vorgeschlagen, die mit den Symbolen I, II, III und IV dargestellt werden, wobei I sich auf das höchste Risiko bezieht und IV auf das schwächste. Auch hier ist eine Anpassung an das jeweilige Einsatzgebiet empfohlen.

Häufigkeit	Konsequenzen				Risiko- klasse	Interpretation
	Katastrophal	Kritisch	Begrenzt	Geringfügig		
Häufig	I	I	I	II	I	nicht tolerierbares, unerträgliches Risiko
Wahrscheinlich	I	I	II	III		
Gelegentlich	I	II	III	III	II	unerwünschtes Risiko, das nur tolerierbar ist, wenn eine Risikominderung nicht durchführbar ist oder die Kosten der Minderung unverhältnismäßig hoch im Vergleich zur erzielten Verbesserung sind
Gering	II	III	III	IV		
Unwahrscheinlich	III	III	IV	IV	III	tolerierbares Risiko, wenn die Kosten einer Risikominderung den Wert der erreichten Verbesserung übersteigen
Unglaublich	IV	IV	IV	IV		
					IV	vernachlässigbares Risiko

Tab. 30: Beispiel einer Risikomatrix aus der DIN EN 61508

Oft werden auch anstelle von Risikoklassen Sicherheits-Integritätslevel (SIL)<sup>56</sup> verwendet [35, 102], welche entsprechend den Risikoklassen eingesetzt werden. Die DIN EN 61511 fordert bei einer vierstufigen Klassifizierung ab SIL = 3 weitere Analysen, da das Risiko nicht ausreichend reduziert ist. Neben der Verwendung von SIL-Werten ist auch eine Verwendung von selbst definierten Risikoklassen möglich [103]. Die tatsächlichen Risikoklassen I, II, III, IV sind daher vom Einsatzgebiet abhängig. Daher sollte diese Tabelle mehr als Praxisbeispiel und weniger als verpflichtende Vorgabe gesehen werden. Eine weitere Möglichkeit der Einstufung ist in Abb. 79 zu erkennen. Hierbei wird deutlich, dass Matrizen auch farblich kodiert werden können, um die Risikoklassifizierung hervorzuheben.

(potentielle) Auswirkung	Eintrittshäufigkeit		
	Selten	Manchmal	Oft
Todesfälle: ernste Auswirkung auf Umwelt/Öffentlichkeit; großer Sachschaden			
Schwere Verletzung; begrenzte Auswirkungen auf Umwelt/Öffentlichkeit; geringer Sachschaden			
Verletzungen; keine Auswirkung auf Umwelt/Öffentlichkeit; geringer Sachschaden			
Unwesentliche Auswirkung auf Mensch/Umwelt/Öffentlichkeit/Sachen			

■ Risiko nicht vertretbar  
■ Risiko nicht ausreichend bestimmt, weitere Analysen erforderlich  
■ Risiko vertretbar

Abb. 79: Beispiel einer Risikomatrix [381]

<sup>56</sup> Sicherheits-Integritätslevel (SIL) (engl. Safety integrity level (SIL)): eine von vier diskreten Stufen, die einem Wertebereich der Sicherheitsintegrität entsprechen, wobei das Sicherheits-Integritätslevel 4 die höchste Stufe der Sicherheitsintegrität und das Sicherheits-Integritätslevel 1 die niedrigste Stufe darstellt [101].

## Risikograph

Die Risikomatrix wird allgemein als ausreichend betrachtet, um einen weiten Anwendungsbereich abzudecken. Es können aber Anwendungen existieren, die Aspekte enthalten, welche die Einführung zusätzlicher Risikoparameter erforderlich machen [102]. Hierzu kann der Risikograph verwendet werden. Bei dem Risikographen wird ein Risiko mit Hilfe einer Anzahl verschiedener Parameter in sinnvoller Kombination eingestuft. Er kann, je nach Einsatzgebiet, in verschiedenen Formen und Ausprägungen angewendet werden [35, 102, 103, 107], ist aber in seiner Grundstruktur immer gleich. Lediglich Parameterauswahl und -anzahl variieren.

Die Methode ist ein qualitatives Verfahren, welches zu abgestuften Risikokategorien oder auch Sicherheits-Integritätsleveln führt. Grundsätzlich basiert dieses Verfahren ebenfalls auf der Gleichung  $R = H \times A$  (Häufigkeit  $\times$  Auftretensschwere). Da die exakte Quantifizierung von Risiken mit dieser Gleichung in bestimmten Fällen nur schwer möglich ist, werden zur Vereinfachung für die Quantifizierung weitere Parameter eingeführt, welche auf den beeinflussenden Faktoren für „H = Häufigkeit“ beruhen. So kann beispielsweise die Häufigkeit eines gefährlichen Vorfalls durch a) die Häufigkeit und Zeit des Aufenthalts im Gefahrenbereich, b) die Möglichkeit, den gefährlichen Vorfall zu vermeiden und c) die Wahrscheinlichkeit des Auftretens des gefährlichen Vorfalls weiter unterteilt werden [102]. Bei Bedarf kann dies noch durch weitere Faktoren ergänzt werden. Bei Beschränkung auf diese Faktoren ergeben sich somit die Parameter:

- Auswirkung des gefährlichen Vorfalls (C),
- Häufigkeit und Zeit des Aufenthalts im Gefahrenbereich (F),
- Möglichkeit, den gefährlichen Vorfall zu vermeiden (P),
- Wahrscheinlichkeit des unerwünschten Ereignisses (W).

Eine Kombination der Parameter ermöglicht die Erstellung eines Risikographen wie in Abb. 80 dargestellt.

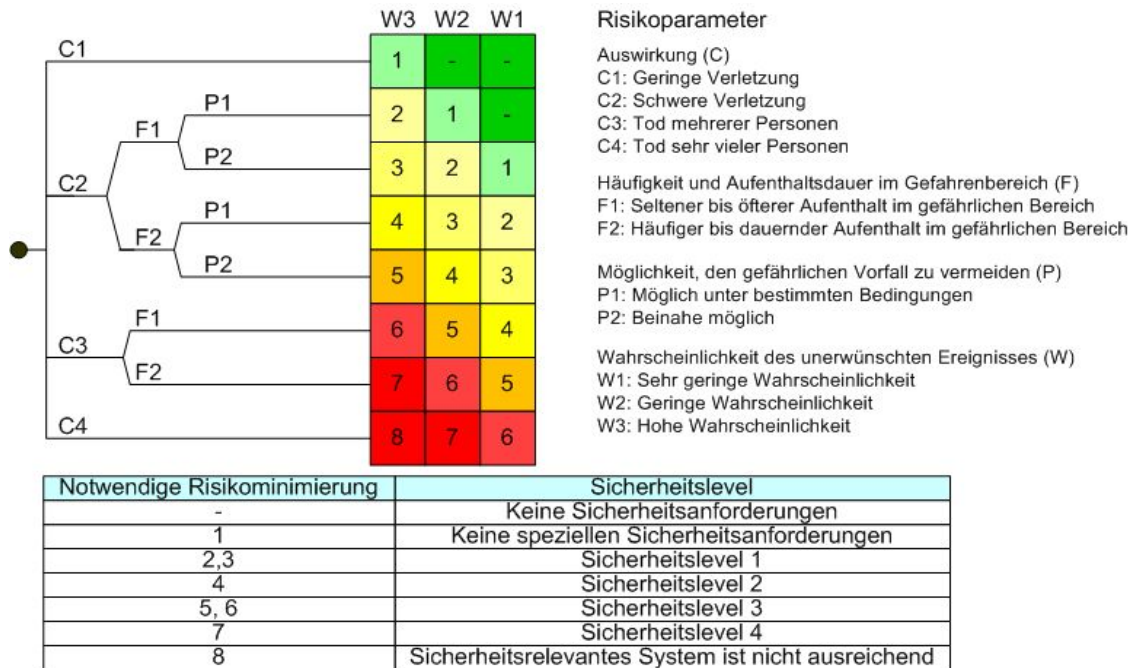


Abb. 80: Allgemeines Prinzip eines Risikographs [adaptiert nach: 102]

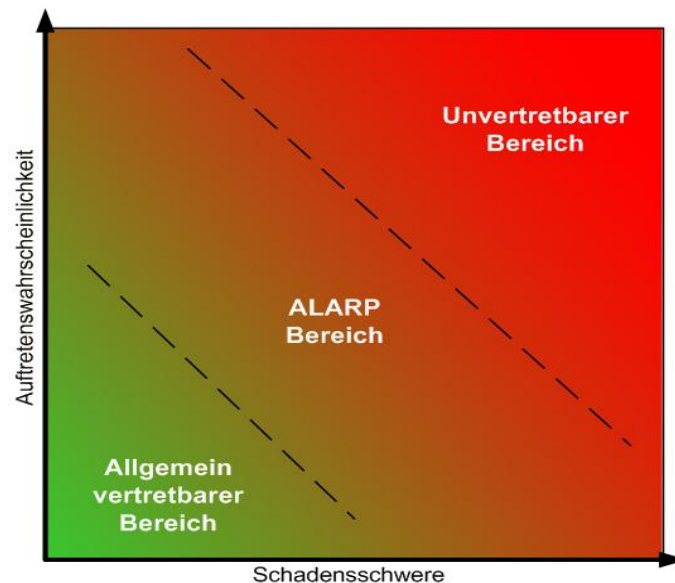
### 5.4.3 Vertretbarkeit des Risikos

Allgemein ist bekannt, dass die Anwendung eines Medizinproduktes einen gewissen Risikograd mit sich bringt. Die Annehmbarkeit des Risikos wird durch die bereits beschriebenen Kriterien „Auftretenshäufigkeit“ und „Schadensausmaß“ beeinflusst. Hinzu kommt noch der mit dem Medizinprodukt verbundene Nutzen, der die Vertretbarkeit eines Risikos bestimmt. Die internationalen Normen und die Literatur legen keine Spezifikationen für die Vertretbarkeit des Risikos fest. Vielmehr erstellt der Hersteller „unter Berücksichtigung des anerkannten Standes der Technik Beurteilungen über die Sicherheit eines Medizinprodukts einschließlich der Akzeptanz von Risiken, um die Eignung eines Medizinprodukts zum Inverkehrbringen für seine Zweckbestimmung festzulegen“ [108]. Somit bestimmt auch der Hersteller die Vertretbarkeit des Risikos. Die hierfür notwendigen Verfahren sind u.a.:

- Anwendung einschlägiger Normen
- Auswertung der Daten aus klinischen Studien und Nutzertests
- Vergleich mit Risikostufen bereits durchgeführter Bewertungen
- Risiko-Nutzen-Analyse

Die Norm DIN EN ISO 14971 legt die Durchführung einer Risikobeurteilung fest, gibt jedoch dazu kein spezielles Verfahren an. Um ein Risiko zu beurteilen, kön-

nen identifizierte Gefährdungssituationen und das damit verbundene Risiko unter Verwendung verfügbarer Informationen oder Daten über Auftretenswahrscheinlichkeit und Schadensausmaß beispielsweise mit Hilfe eines Risikodiagramms eingeschätzt werden (Abb. 81) [102, 108]. Dieses kann zur Definition eines Risiko-Niveaus herangezogen werden. Dabei wird berücksichtigt, dass in einigen Fällen ein Risiko, angesichts des mit dem Medizinprodukt verbundenen Nutzens, eher vertretbar ist, als in anderen Fällen.



**Abb. 81: Beispiel eines Risikodiagramms mit drei Bereichen in Anlehnung an die DIN EN ISO 14971 [106]**

Bei dem Risikodiagramm wird aus den Faktoren Eintretenswahrscheinlichkeit und Schadensausmaß der ermittelte Wertebereich in drei Bereiche eingeteilt. Der allgemein vertretbare Bereich inkludiert Risiken, die so niedrig sind, dass sie im Vergleich mit anderen Risiken und angesichts des Nutzens der Anwendung des Medizinprodukts vernachlässigbar sind. Dem unvertretbaren Bereich werden Risiken zugeordnet, die aufgrund ihrer Auftretenswahrscheinlichkeit und Schadensschwere als inakzeptabel angesehen werden können. Sie sind so groß, dass sie insgesamt abgelehnt werden müssen. Es kann aber auch sein, dass ein Risiko zwischen den beiden Bereichen liegt und so weit verringert ist, dass es zwar nicht dem allgemein vertretbaren Bereich zugeordnet werden kann, aber auf die niedrigste praktikable Stufe gesenkt wird. Hierbei greift das ALARP-Prinzip (as low as reasonably practicable). Das Risiko ist so niedrig, wie vernünftigerweise möglich, und somit, angesichts des gestifteten Nutzens, tolerierbar. Der tolerierbare Bereich unterscheidet sich hierbei von dem vertretbaren Bereich dadurch,

dass hier die Bereitswilligkeit angezeigt wird, mit einem Risiko unter Berücksichtigung des gestifteten Nutzens zu leben [102, 108].

Um Risiken dahingehend zu überprüfen, ob sie nach Erschöpfung aller risikomindernder Maßnahmen vom inakzeptablen Bereich in den ALARP-Bereich übertragen werden können, kann eine Risiko-Nutzen-Analyse durchgeführt werden. Diese Methode eignet sich für die Überprüfung des Nutzens und der Schadenswahrscheinlichkeit. Hierzu wird der gestiftete Nutzen dem Risiko gegenübergestellt. Eine ausführliche Beschreibung liefert die DIN EN ISO 14971 [108]. Nach Bewertung der Risiken als akzeptabel oder vertretbar, ist die Risikoanalyse abgeschlossen, und es kann von einer Sicherheit ausgegangen werden, entsprechend der bereits genannten Definition von der „Freiheit von nicht akzeptablen Risiken“.

## **5.5 Die Komponente Mensch im Risikomanagement**

Gefahren sind oft Ergebnisse einer falschen Geräteverwendung (vgl. Kap. 5.1.3). Diese Gefahren, welche speziell auf die Mensch-Maschine-Interaktion zurückzuführen sind, werden gebrauchsbazogene Gefährdungen (engl. use related hazard) genannt. Nach Kaye & Crowley werden gebrauchsbazogene Gefährdungen durch einen oder mehrere der folgenden Gründe verursacht [212]:

- Geräte werden auf eine nicht vorhersehbare Weise verwendet,
- Geräte werden auf eine erwartete Weise verwendet, jedoch wurde diese nicht ausreichend überprüft,
- Die Gerätebedienung ist inkonsistent bzgl. Benutzererwartungen oder Intuitionen,
- Die Nutzungsumgebung beeinflusst die Gerätebedienung,
- Die körperlichen, wahrnehmungsspezifischen oder kognitiven Kapazitäten werden bei der Gerätebedienung unter gewissen Umständen überschritten.

Risikomanagement alleine ist hierbei nicht länger ausreichend. Durch die Erkenntnis, dass immer mehr Risiken durch die Mensch-Maschine-Interaktion begründet sind, wird ersichtlich, dass der menschliche Einfluss bei der Risikobetrachtung mit berücksichtigt werden muss. Zwar wird in den europäischen und nationalen Normen des Risikomanagements [94, 108] der Einfluss des Benutzers am Rande berücksichtigt, jedoch werden ihm weder die nötige Bedeutung zugewiesen, noch werden daraus resultierende explizite Handlungsschritte erörtert. Eine genauere Betrachtung erfolgte in der Medizin erst durch die Normen DIN EN



IEC 62366 und DIN EN IEC 60601-1-6 (vgl. Kap. 2.3). Auch die US-amerikanische „Food and Drug Administration“ (FDA) verlangt von den Herstellern, menschliche Einflüsse und ergonomische Gesichtspunkte (vgl. Kap. 2) in der Risikobetrachtung mit zu berücksichtigen. Zu diesem Zweck wurde ein Leitfaden entwickelt, welcher diese Gesichtspunkte in den Risikomanagementprozess mit einbezieht [212, 396].

Dennoch kann die Schätzung des Risikos einer Gefahr, welche von einer Gerätebedienung ausgeht, schwierig und irreführend sein [212]. Probleme der Mensch-Maschine-Interaktion, welche zu gefährlichen Situationen führen können, sind schwer vorhersehbar, da es viele verschiedene Möglichkeiten und Umstände gibt, in denen Benutzer und das Gerät interagieren. So können einige gebrauchsbetonte Gefährdungen nicht beim Risikomanagement betrachtet werden. Auch wenn Gefahren erwartet werden, ist es schwierig, ihre Auftretenswahrscheinlichkeit vorherzusagen [212]. Selbst Benutzerstudien können irreführende Ergebnisse der Wahrscheinlichkeit liefern, da die beobachteten gebrauchsbetonten Gefährdungen eher geringer sind als bei der wirklichen Verwendung. Deshalb ist bei einer Benutzerstudie so weit wie möglich der Nutzungskontext nachzuempfinden und, wenn notwendig, eine Stresssituation zu erzeugen. Bei medizinischen Geräten kann ein Risiko, welches mit der geringen Wahrscheinlichkeit einer gebrauchsbetonten Gefährdung verbunden ist, irreführend sein, da Benutzer oft nicht in der Lage sind, in selten auftretenden Situationen adäquat zu reagieren. Daher ist innerhalb des Risikomanagementprozesses bei Bedienrisiken eine sorgfältige Prüfung der Schwere des Schadens besonders wichtig. Diese Prüfung sollte folgende Aufgaben beinhalten [212]:

1. Identifikation und Beschreibung benutzungsbedingter Gefahren durch eine Analyse der bestehenden Informationen,
2. Anwendung empirischer Methoden unter Mitwirkung repräsentativer Geräteanwender, um Gefahren zu identifizieren und zu beschreiben, welche sich nicht durch analytische Methoden erkennen bzw. verstehen lassen,
3. Einschätzung des Risikos eines jeden gebrauchsbetonten Gefährdungsszenarios,
4. Entwicklung von Strategien und Kontrollen zur Verringerung der Auftretenswahrscheinlichkeit oder Milderung der Folgen von gebrauchsbetonten Gefährdungen,
5. Auswahl und Anwendung von Kontrollstrategien,

6. Sicherstellung der Angemessenheit und Effektivität von Kontrollen der Risikoverringerung,
7. Ermittlung potentieller neuer Gefahren als Folgen der durchgeführten Kontrollstrategie,
8. Sicherstellung funktionaler und operativer Anforderungen an das System,
9. Validierung der sicheren und effektiven Gerätebedienung.

Von hoher Wichtigkeit beim Verständnis gebrauchbezogener Gefährdungen ist ein genaues und vollständiges Wissen über die Gerätebedienung. Kenntnisse über die Verwendung von Technologie durch den Menschen, resp. Wahrnehmung der hier stattfindenden Interaktionen und deren Optimierung, sind Bestandteile der Themenbereiche Usability Engineering und Mensch-Maschine-Kommunikation (vgl. Kap. 2).

Im vorherigen Kapitel wurde beschrieben, wie im Rahmen der Gebrauchstauglichkeitsevaluation Anwendungsszenarien aus dem Nutzungskontext identifiziert werden, die zu gefährlichen Situationen führen können (vgl. Kap. 4). Abb. 82 zeigt die Struktur eines derartigen Anwendungsszenarios. Verdeutlicht wird die Abhängigkeit der Geräteverwendung von der Nutzungsumgebung sowie von den Nutzer- und Gerätecharakteristika. Sobald diese Einflussgrößen in einer Anforderungsanalyse (vgl. Kap. 4.1.1) identifiziert sind, können sie als Ursachen oder beeinflussende Faktoren von gebrauchbezogenen Gefährdungen untersucht und in einer Gebrauchstauglichkeitsevaluation verifiziert werden. Sie können damit einer Bewertung der Gebrauchstauglichkeit des medizinischen Gerätes dienen oder im Rahmen des Usability Engineerings nützlich sein, um Probleme in der Mensch-Maschine-Interaktion zu verstehen und zu vermeiden oder risikomindernde Strategien zu entwickeln.

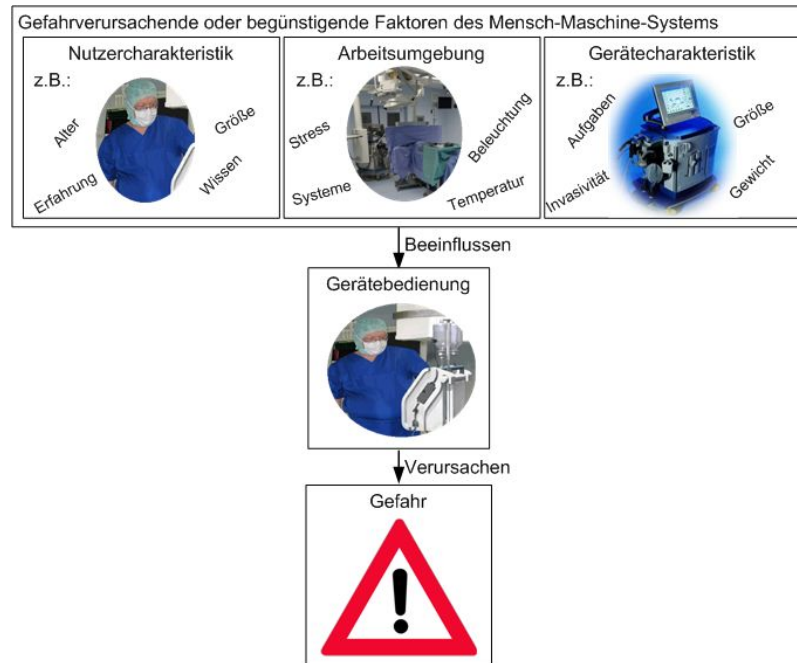


Abb. 82: Abhängigkeit der Geräteanwendung vom Nutzungskontext in der Medizin [212]

Grundsätzlich ähnelt der Risikoprozess unter besonderer Berücksichtigung der gebrauchsbazogenen Gefährdungen (Abb. 83) dem „normalen“, bereits beschriebenen Risikoprozess (vgl. Kap. 5.4). Empirische Methoden, wie z.B. eine Gebrauchstauglichkeitsuntersuchung, können in fast allen Phasen angewendet werden [212].

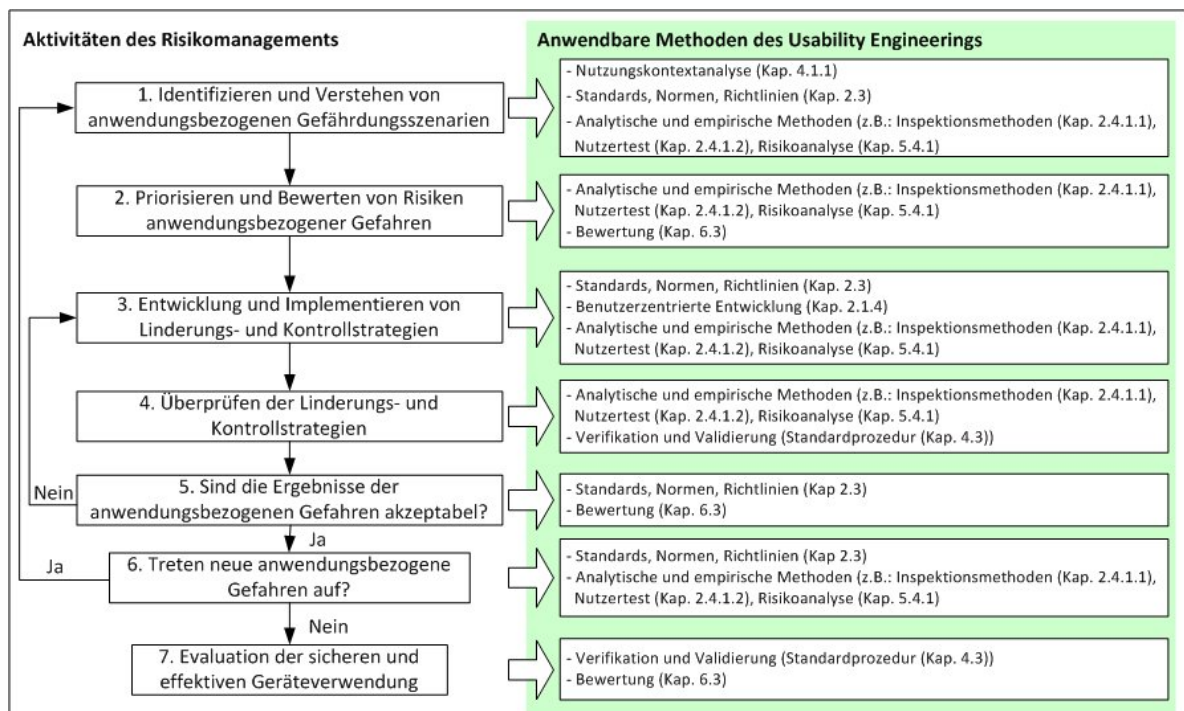


Abb. 83: Methoden des Usability Engineerings im Risikomanagement [adaptiert nach 212]

Durch die Betrachtung von Usability Engineering in Kombination mit Risikomanagement-Prozessen sind Hersteller nicht nur auf die Zweckbestimmung des medizinischen Gerätes (engl.: device intended use) fokussiert, sondern betrachten auch die Benutzerinteraktion mit dem Gerät [346]. Eine Gegenüberstellung des traditionellen Risikomanagements und des Risikomanagements unter Beachtung menschlicher Faktoren (Human Factors Analysis) liefert Patterson (Tab. 31):

<b>Traditionelles Risikomanagement</b>	<b>Menschl. Faktoren im Risikomanagement</b>
Fokus auf das Gerät	Fokus auf den Nutzer und die Interaktion mit dem Gerät
Betrachtung der Kontrolle von Hard- und Software sowie mechanischer und elektronischer Fehlerquellen	Untersuchung von Nutzerfehlern während der Verwendung von Geräten beim Wahrnehmen, Verstehen und Ausführen
Zusammentragen von Fehlerszenarien unter Abwesenheit der repräsentativen Nutzer	Genaue empirische Evaluation der tatsächlichen Mensch-Maschine-Interaktion
Blockschaltbilder zeigen Ein- und Ausgaben der einzelnen Systemkomponenten, welche sich auf die vorhersehbare Gerätereaktion beziehen	Kontinuierliche und schrittweise Analyse zur Identifikation von Nutzeranforderungen und potentiellen Fehlern
Fehlermöglichkeits- und Einflussanalyse (FMEA) ist auf das Gerät fokussiert	Fehleranalyse ist auf den Menschen fokussiert

**Tab. 31: Traditionelles Risikomanagement vs. Human Factors Analysis [306]**

## 5.6 Sicherheit und Gebrauchstauglichkeit

In sicherheitskritischen Systemen ist die Gebrauchstauglichkeit von besonderer Bedeutung, weil durch Defizite in diesem Bereich sicherheitsbedrohende Fehler entstehen können [140, 146, 332]. Gerade bei solchen Systemen scheinen nach Fields et al. zwei widersprüchliche Meinungen zu existieren. Zum einen wird vielfach angenommen, dass fortschreitende Technologien und Automatisierungen ineffizienter und fehlerträchtiger Aufgaben dazu führen könnten, dass Leistung und Sicherheit verbessert würden, zum anderen wird die Meinung vertreten, dass die Einführung neuer Technologien in immer komplexer werdende Systeme negative und unerwartete Folgen haben kann. Die Fähigkeit des Menschen, effektiv arbeiten zu können, ist gerade bei sicherheitskritischen Systemen, zu denen das medizinische Arbeitssystem gehört, von immenser Bedeutung [140].

Die Expertengruppe des Computer-Human-Interaction Workshops CHI 98 [300] stellte bereits 1998 fest, dass durch Gebrauchstauglichkeitsprobleme hervorgerufene Fehler in sicherheitskritischen Systemen infolge vermehrten Technikeinsatzes zunehmend an Bedeutung gewinnen werden. Die Realität zeigt, dass viele Unfälle, auch in der Medizin, aus Fehlern in der Mensch-Maschine-Interaktion resultieren (vgl. Kap. 2.2.1.), was für die Richtigkeit der CHI-Expertenmeinung

spricht. Problematisch ist jedoch, dass bei der Gestaltung von sicherheitskritischen Systemen Gebrauchstauglichkeits- und Sicherheitsprobleme getrennt voneinander betrachtet werden [304]. Die Methoden der Gebrauchstauglichkeits-evaluation (vgl. Kap. 2.4.1) sind grundsätzlich nützlich, um Probleme der Benutzerschnittstelle aufzudecken, schenken aber dem Thema Sicherheit keine besondere Beachtung. Spezielle Techniken zur Aufdeckung von Sicherheitsproblemen, wie z.B. HAZOP und FMEA betrachten nicht das Design und die Benutzerschnittstelle (vgl. Kap. 5.4.1). Wichtig ist, dass beide Wissensgebiete verbunden werden, und dass in sicherheitskritischen Systemen die Gebrauchstauglichkeit zusammen mit der Sicherheit betrachtet wird.

Ein weiteres großes Problem bei Benutzerschnittstellen sicherheitskritischer Geräte in der Medizin ist der Faktor der fehlenden Konsistenz. Nach Basnyat et al. kann durch Konsistenz die Zuverlässigkeit, Leistungsfähigkeit, Gebrauchstauglichkeit und Fehlertoleranz eines Systems verbessert werden [15]. Die Ursache für diese Problematik in der Medizin ist darin zu suchen, dass verschiedene Hersteller, resp. deren Entwicklungsabteilungen, unterschiedliche Sichtweisen für Bedienkonzepte und das Design haben. Das medizinische Arbeitssystem ist geprägt durch eine Vielzahl medizinischer Geräte unterschiedlichen Typs von unterschiedlichen Herstellern. Daher wird der Nutzer im medizinischen Arbeitssystem auch mit einer Vielzahl unterschiedlicher Bedienkonzepte und Schnittstellen konfrontiert. Durch deren Vereinheitlichung könnten konsistente Bedingungen geschaffen und Fehler verringert werden [15]. Um dies zu erreichen, müssten jedoch Modelle entwickelt werden, die ein Mindestmaß an Konsistenz sicherstellen. Arbeiten, wie z.B. die von Kunert [224] haben bereits für andere Systeme Vorschub geleistet. Bei diesen Arbeiten werden Gestaltungslösungen für spezielle Interaktionsschritte vorgeschlagen, welche sich als gebrauchstauglich bewährt haben. Bei sicherheitskritischen Systemen sollte jedoch nicht nur die Fehlerprävention beachtet werden, sondern auch begangene Fehler. Diese müssen, sobald ein Problem aufgetreten ist, diagnostiziert werden, unabhängig davon, ob ein Schaden entstanden ist oder nicht. Fehlerauswirkungen müssen eingedämmt und Maßnahmen geschaffen werden, um ein System schnell wieder in einen sicheren Zustand zu bringen [146].

## 5.7 Menschliche Fehler und Sicherheit

Menschen, die sicherheitskritische Geräte bedienen, sind Teil eines sicherheitskritischen Systems (vgl. Kap. 5.3.1). Menschliches Verhalten ist in der Regel unvorhersehbar und unzuverlässig. In manchen Bereichen werden daher Menschen durch Maschinen ersetzt. Trotzdem sind sie aus den meisten Systemen nicht wegzudenken, da sie wesentlich flexibler sind und besser auf unerwartete Situationen reagieren können als Maschinen. Dennoch wird der Mensch vielfach, wenn auch zu unrecht, als das schwächste Glied eines Mensch-Maschine-Systems angesehen, da eine große Zahl aller vorkommenden Unfälle auf menschliches „Fehlverhalten“ zurückgeführt wird [220]. Richtigerweise ist eher von einem „Interaktions- und Kommunikationsversagen des Mensch-Maschine-Systems“ zu sprechen [180], wenn auch erkenntlich ist, dass die Komponente Mensch in den meisten Fällen der auslösende Teil des Fehlers ist, weil er eine falsche Handlung ausübt oder eine richtige unterlässt, unabhängig davon, ob das Problem von ihm selbst oder von der Maschine verursacht wird. Es ist davon auszugehen, dass die Fehlerrate von Art und Umfang der Tätigkeit sowie von Umwelteinflüssen stark beeinflusst wird (vgl. Kap. 1.2.2.1) [309]. Zwar können Menschen im Sinne eines sicherheitsgemäßen Verhaltens durch Schulung, Motivation etc. beeinflusst werden, doch weist Rohmert andererseits nach, dass selbst hoch motivierte und gut bezahlte Versuchspersonen nicht fehlerfrei arbeiten können [342]. Externe Einflüsse fördern das Eintreten eines Fehlers zusätzlich. Peters & Meyna zeigen, dass beispielsweise unter Zeitstress die Fehlerwahrscheinlichkeit zunimmt. Bei Unterlassungsfehlern lässt sich hierbei ein linearer Verlauf feststellen [309].

Die Zuverlässigkeit eines Menschen lässt sich analog zu der Zuverlässigkeit eines technischen Systems mit der Wahrscheinlichkeit definieren, dass der Mensch während eines vorgegebenen Zeitintervalls unter vorgegebenen Bedingungen diejenigen Funktionen erfüllt, für die er vorgesehen ist [309]. Diese Definition impliziert auch die fehlerfreie Erfüllung seiner Aufgaben. Fehlhandlungswahrscheinlichkeiten für menschliche Handlungsfehler bei unterschiedlichen Aufgaben lassen sich nach Watson [407] unter Verwendung von Reasons Taxonomie (vgl. Kap. 5.1.2) statistisch wie folgt angeben:

- fähigkeitsbasiert (skill-based):  $5 \cdot 10^{-5}$  bis  $5 \cdot 10^{-3}$
- regelbasiert (rule based):  $5 \cdot 10^{-4}$  bis  $5 \cdot 10^{-2}$
- wissensbasiert (knowledge based):  $5 \cdot 10^{-3}$  bis  $5 \cdot 10^{-1}$

Hierbei wird erkenntlich, dass Fehlhandlungswahrscheinlichkeiten mit zunehmender Automatisierung der Tätigkeit abnehmen [409]. Unter Einbeziehung der statistischen Wahrscheinlichkeiten für einen menschlichen Fehler, der Individual- und Situationsfaktoren und der Wahrscheinlichkeit eines resultierenden Vorfalls könnte somit auf die Wahrscheinlichkeit eines unerwünschten Ereignisses geschlossen werden. Spezifische Berechnungsmodelle unter Berücksichtigung externer Einflüsse, wie z.B. Training, Erfahrungen und Stress sind sehr mannigfaltig und konnten bisher noch nicht in ihrer vollen Komplexität erfasst werden [180]. Lediglich Berechnungen mit einer Variablen aber ansonsten konstanten Bedingungen (Ceteris-paribus-Annahme) sind genauer erforscht [41].

## **5.8 Fehlerbewertung in der Mensch-Maschine-Interaktion**


Eine Bewertung ist ein Vorgang der Zuweisung eines wertenden Prädikats und kann sowohl zur Qualitätsbeurteilung als auch zur Entscheidungsfindung herangezogen werden. Für die Bewertung von Gebrauchstauglichkeitsproblemen gibt es bereits verschiedene Ansätze [11, 72, 108, 124, 125, 284, 348]. Diese sind jedoch nicht auf sicherheitskritische Systeme wie das medizinische Arbeitssystem ausgerichtet. Halbeck et al. fordern ein einheitliches Bewertungssystem, welches die unterschiedlichen Anforderungen an die medizinische Gebrauchstauglichkeit erfüllt und als Grundlage für Zertifizierer (z.B. TÜV, FDA) dienen kann [14].

Um Gebrauchstauglichkeit medizinischer Geräte zu bewerten, müssen die hieraus entstehenden Fehler und deren Ursache eingeschätzt werden. Dementsprechend wird im Folgenden zuerst auf die Fehleridentifikation eingegangen, um danach die Fehler und die Problembewertung zu veranschaulichen. Hierzu werden geeignete Bewertungsmethoden aus dem Bereich des Risikomanagements und dem Bereich der Gebrauchstauglichkeitsbetrachtung genauer erläutert. Fehler können entsprechend den bereits betrachteten Grundlagen nach Auftretenswahrscheinlichkeit, Schweregrad und potentiellem Schaden bewertet werden.

### 5.8.1 Fehleridentifikation

*Benutzerprobleme treten nicht dort auf, wo man sie erwartet;  
Bereiche, die perfekt erscheinen, erweisen sich oft als problematisch*  
B. Tognazzini

Erster wichtiger Schritt in der Fehlerbeurteilung ist die Identifikation des Fehlerausgangspunktes. Es muss eingeschätzt werden, ob es sich um einen Maschinenfehler, um menschliches Versagen, einen Fehler in der Mensch-Maschine-Kommunikation oder um eine Kombination dieser Ursachen handelt (Kap. 5.2). Hierzu müssen die Fehlerquelle und die für den Fehler verantwortliche Komponente bzw. Komponentenkombination identifiziert werden [348], um dann auf mögliche Gebrauchstauglichkeitsprobleme als Fehlerursache schließen zu können. Diese Prozedur kann einfach sein, ist aber in vielen Fällen äußerst schwierig, da Fehler oft durch komplexe Wirkungszusammenhänge wie beispielsweise zwischen Wahrnehmung, Informationsverarbeitung der Geräte und Umwelteinflüssen entstehen können. Um eine Fehleranalyse durchzuführen, müssen meist viele Bereiche betrachtet werden. Hilfreich sind hierbei Testprotokolle, Video- und Audiodokumentationen sowie ein ausreichendes Verständnis für die komplexen Vorgänge der Mensch-Maschine-Interaktion [348]. In Tab. 32 werden Beispiele beschrieben, in denen von dem Fehler auf die Fehlerquelle geschlossen wird.

Aufgabe	Fehlerquelle
Spezifizieren Sie das Laufwerk, in welchem das Programm nach der zu konfigurierenden Applikation suchen soll.	<ol style="list-style-type: none"> <li>1. Die Sprache auf dem Bildschirm war zu technisch. Der Nutzer hat das Icon für das Laufwerk nicht bemerkt. Das Wissen des Nutzers wurde überschätzt.</li> <li>2. Die Bedienungsanleitung hatte keine Beschreibung, wie das Programm ein Laufwerk sucht.</li> </ol>
Verwenden Sie die Hilfefunktion, um zusätzliche Informationen zu erhalten.	<p>Das Fragezeichen am Bildschirm oben rechts wird nicht erkannt (Abb. links). Es ist zu exponiert und hat keine Verbindung zum Hauptbildschirm. Nutzer übersehen es, da kein Bezug zum Bedienkonzept registriert wird. Durch eine direkte Verbindung mit dem Screen wird dieses Problem behoben (Abb. rechts).</p> 
Bestücken Sie die Spritzenpumpe mit einer 50ml Spritze.	<ol style="list-style-type: none"> <li>1. Die Spritze kann in verschiedenen Positionen eingelegt werden.</li> <li>2. Die Erklärung ist nicht verständlich, Bilder sind nicht klar genug.</li> <li>3. Der Spritzenbügel hat kein Icon.</li> </ol>

Tab. 32: Beispiele für die Fehlersuche



### **5.8.2 Bewertung der Bedeutung eines Fehlers**

Um einen Fehler zu bewerten, muss seine Bedeutung erfasst werden. Meist wird dies von mehreren Experten durchgeführt, welche die aufgedeckten Probleme nach ihrer jeweiligen Schwere bewerten [284]. Bereits Dumas & Redish stellten sich die Frage, was die Bedeutung eines Problems ausmacht [124]. Für sie galten die Dimensionen „scope – How widespread is the problem?“ (Umfang - Wie verbreitet ist das Problem?) und „severity – How critical is the problem?“ (Schweregrad - Wie kritisch ist das Problem?) als Maßstab. Weitere Schemata mit unterschiedlichem Vokabular aber gleichem Inhalt werden auch von Nielsen, Desurvire, Rubin und Hollingsed et al. angewendet [72, 184, 284, 348]. Rubin verwendet beispielsweise den kritischen Aspekt, welcher das Zusammenwirken von Schweregrad (severity) und Auftretenswahrscheinlichkeit (probability of occurrence) darstellt [348]. Nielsen unterteilt die Gewichtung der Gebrauchstauglichkeitsprobleme nach der Auftretensfrequenz (frequency), der Bedeutung (Ist das Problem schwierig oder einfach zu bewältigen?) und der Dauer (Ist es ein einmaliges Problem oder werden Nutzer wiederholt von ihm gestört?) [289]. In sicherheitskritischen Systemen und somit auch in der Medizintechnik sollte zusätzlich noch eine Bewertung nach der Auswirkung stattfinden. Im Folgenden werden verschiedene Bewertungsmethoden für die einzelnen Größen vorgestellt, um letztendlich eine Bewertungsmethode für den medizinischen Bereich zu entwickeln.

### **5.8.3 Beurteilung nach Auftretenswahrscheinlichkeit**

Nach DIN EN ISO 14971 ist die Auftretenswahrscheinlichkeit der „Parameter des Risikos, der Auskunft über die Wahrscheinlichkeit gibt, mit der eine identifizierte Gefährdung bzw. ihre Ursache in der Praxis tatsächlich auftreten könnte“. Die Kategorien ihrer Bewertung sind z.B. häufig, wahrscheinlich, gelegentlich, entfernt, vorstellbar, unwahrscheinlich und unvorstellbar [108]. Diese Definition gilt zwar für Risiken, kann jedoch auch für die Auftretenswahrscheinlichkeit von Gebrauchstauglichkeitsproblemen verwendet werden. Nielsen unterscheidet, ob das Problem bei einigen oder vielen Nutzern auftritt [284]. Rubin hingegen schlägt eine Schätzung der Auftretensfrequenz im Feld vor. Um diese Schätzung zu erhalten, berechnet er das Produkt aus der Prozentzahl der betroffenen Nutzer und der Wahrscheinlichkeit, mit der die betroffene Gruppe mit dem Problem konfrontiert wird. Werden beispielsweise 10% der Population in 50% der Fälle kon-

frontiert, dann liegt die Auftretenswahrscheinlichkeit bei 5% ( $0,1 \times 0,5 = 0,05$ ) [348]. In einem folgenden Schritt können den entsprechenden Prozentwerten Frequenzränge zugewiesen werden (Tab. 33).

Frequenzranking	Geschätzte Auftretensfrequenz
4	Wird in $\geq 90\%$ der Zeit auftreten, in der das Produkt verwendet wird.
3	Wird in 51-89% der Zeit auftreten, in der das Produkt verwendet wird.
2	Wird in 11-50% der Zeit auftreten, in der das Produkt verwendet wird.
1	Wird in $\leq 10\%$ der Zeit auftreten, in der das Produkt verwendet wird.

**Tab. 33: Ranking der Auftretenswahrscheinlichkeit eines Gebrauchstauglichkeitsproblems**

Dumas & Redish unterscheiden zusätzlich zwischen lokalen und globalen Problemen. Normalerweise werden Probleme umso wichtiger, je globaler sie verbreitet sind. Globale Probleme werden allerdings von vielen Personen mit geringer Erfahrung und Training in der Mensch-Maschine-Interaktion selten als solche erkannt, sondern eher als eine Vielzahl verschiedener lokaler Probleme. Sie haben daher auch meist eine größere Bedeutung als lokale Probleme [124].

#### 5.8.4 Beurteilung nach dem Schweregrad

Eine Beurteilung des Problems nach dem Schweregrad ist sinnvoll, um es zu erfassen und zu gewichten. Beispielsweise gibt es Probleme, die lediglich den Nutzer in seiner Zufriedenheit beeinflussen, aber auf die Effizienz und Effektivität der Aufgabenerfüllung keine Auswirkungen haben. Es gibt aber auch Probleme, die dazu führen, dass der Nutzer seine Arbeitsaufgabe nicht weiter erledigen kann. In der Praxis werden meist 3-4-stufige Problemdifferenzierungen verwendet, um Gebrauchstauglichkeitsprobleme nach ihrer Priorität zu sortieren [11, 124, 283, 284, 348]. In den folgenden Tabellen werden verschiedene Einstufungsskalen vorgestellt (Tab. 34 - Tab. 36).

Ranking	Beschreibung	Definition
4	Unbenutzbar	Die Nutzung ist nicht möglich, oder der Nutzer verweigert die Nutzung eines bestimmten Teils des Produktes aufgrund des Designs oder der Implementierung.
3	Schwerwiegend	Der Nutzer würde gerne das Produkt nutzen, aber er wird erheblich daran gehindert. Der Nutzer wird Schwierigkeiten haben das Problem zu umgehen.
2	Mittelmäßig	Der Nutzer wird in den meisten Fällen das Produkt benutzen können, wird aber einen gewissen Aufwand haben um die Probleme zu umgehen.
1	Irritierend	Das Problem tritt sporadisch auf und kann einfach umgangen werden, oder es ist abhängig von einem Standard, der außerhalb der Produktgrenzen liegt. Es kann auch ein rein kosmetisches Problem sein.

**Tab. 34: Vier-Ebenen-Ranking zur Bewertung des Schweregrades von Gebrauchstauglichkeitsproblemen nach Rubin [348]**




Ranking	Beschreibung	Definition
0	Kein Gebrauchstauglichkeitsproblem	Das Problem lässt sich nicht auf ein Gebrauchstauglichkeitsproblem zurückführen.
1	Kosmetisches Problem	Probleme, die leicht störend wirken, aber die Gebrauchstauglichkeit, wenn überhaupt, nur gering beeinflussen (z.B. eine geringe Bildschirmauflösung, die die Gebrauchstauglichkeit nicht beeinflusst, aber von einer Testperson bemängelt wird)
2	Geringfügiges Gebrauchstauglichkeitsproblem	Das Problem hat nur geringen Einfluss auf die Gebrauchstauglichkeit. Dazu gehören beispielsweise Probleme, die sich durch einen einmaligen Lernprozess ausgleichen lassen.
3	Großes Gebrauchstauglichkeitsproblem	Das Problem beeinflusst die Gebrauchstauglichkeit stark, wie beispielsweise durch nicht selbstbeschreibungsfähige Verweise oder eine undurchsichtige Struktur.
4	Gebrauchstauglichkeitskatastrophe	Probleme, die das Arbeiten nahezu unmöglich machen. Hierzu gehören beispielsweise nicht funktionierende oder falsche Verweise.

**Tab. 35: Fünf-Ebenen-Ranking zur Bewertung des Schweregrades von Gebrauchstauglichkeitsproblemen nach Nielsen [284]**

Level	Beschreibung	Beispiel
Level 1	Problem hindert an der Aufgabenerfüllung	Nutzer geben nach drei Versuchen auf, einen Brief zu bedrucken, da sie nicht wissen, wie sie den Briefbogen in den Drucker einführen müssen.
Level 2	Problem erzeugt signifikante Verzögerung und Frustration	Fehlendes Feedback über einen erfolgreichen Aufgabenabschluss veranlasst den Nutzer, die Aufgabe ein weiteres Mal zu erledigen, um sicher zu gehen, dass die Aufgabe richtig erfüllt wurde.
Level 3	Problem hat geringfügigen Einfluss auf die Gebrauchstauglichkeit	Die Verwendung desselben Schlagwortes für zwei verschiedene Aktionen führt zu kurzfristiger Irritation.
Level 4	Subtiles Problem, welches bei einer späteren Erweiterung behoben werden kann	Ein Nutzer beschwert sich über das antiquierte Design

**Tab. 36: Vier-Ebenen-Ranking zur Bewertung des Schweregrades von Gebrauchstauglichkeitsproblemen nach Dumas [124]**

Für die Gebrauchstauglichkeitsbewertung im medizinischen Bereich wurde von Backhaus in Anlehnung an die DIN EN 614-1 [99] ein Ampelschema mit drei Bewertungsstufen verwendet (Tab. 37) [11].

	Schweres Bediendefizit: Maßnahmen erforderlich	Diese Zone ist gekennzeichnet durch: <ul style="list-style-type: none"> <li>- ergonomische Grundsätze sind nicht erfüllt</li> <li>- Bedingungen, die zu einem unsicheren Betrieb führen können</li> </ul>
	Potentielles Bediendefizit: Maßnahmen sinnvoll	Ergonomische Grundsätze sind erfüllt bei Aufgaben <ul style="list-style-type: none"> <li>- mit zeitlich begrenzter Nutzung</li> <li>- von kurzer Dauer</li> </ul>
	Kein Defizit: keine Maßnahmen erforderlich	<ul style="list-style-type: none"> <li>- sichere Konstruktion</li> <li>- sicherer Betrieb</li> <li>- ergonomische Grundsätze sind erfüllt bei <ul style="list-style-type: none"> <li>▪ häufig genutzten Aufgaben</li> <li>▪ länger andauernden Aufgaben</li> <li>▪ Aufgaben mit Komfort (Wohlbefinden)</li> </ul> </li> </ul>

**Tab. 37: Drei-Ebenen-Ranking zur Bewertung des Schweregrades von Gebrauchstauglichkeitsproblemen nach Backhaus [11]**

### 5.8.5 Klassifizierung nach potentiellm Schaden

Für auftretenden Schaden werden in der Literatur verschiedene Bewertungssysteme vorgeschlagen. Industriebereiche, welche auf dem Gebiet der Risikobewertung erfahrener als der medizinische Bereich sind, haben sich hier auf spezielle Definitionen geeinigt. So lässt sich für den Flugverkehr ein fünfstufiges Bewertungsverfahren für die Situation eines Ausfalls finden (Tab. 38) [35].

Stufe	Folgen eines Ausfalls:
A	<b>Katastrophal:</b> Ein Ausfall verhindert die sichere Fortsetzung des Fluges bzw. die sichere Landung. Der Ausfall bedeutet den Verlust von Menschenleben.
B	<b>Gefährlich/ernst:</b> Ein Ausfall schränkt die Fähigkeit des Flugzeugs oder der Besatzung ein, mit den widrigen Betriebsbedingungen fertig zu werden. Mögliche Konsequenzen sind eine erhebliche Verringerung der Sicherheitstoleranzen oder der Funktionsfähigkeit, eine physische Beeinträchtigung oder eine erhöhte Arbeitsbelastung der Besatzung, die sie möglicherweise an der zuverlässigen oder vollständigen Erledigung ihrer Aufgabe hindert oder negative Auswirkungen auf die Insassen hat (mit ernstesten oder potentiell tödlichen Verletzungen für eine kleine Anzahl dieser Insassen).
C	<b>Erheblich:</b> Ein Ausfall schränkt die Fähigkeit des Flugzeugs oder der Besatzung ein, mit den widrigen Betriebsbedingungen fertig zu werden. Mögliche Konsequenzen sind eine erhebliche Verringerung der Sicherheitstoleranzen oder der Funktionsfähigkeit, eine erheblich erhöhte Arbeitsbelastung für die Besatzung oder Bedingungen, die die Effizienz der Besatzung beeinträchtigen sowie Widrigkeiten für die Insassen (unter Umständen auch Verletzungen).
D	<b>Geringfügig:</b> Ein Ausfall auf dieser Ebene ist für die Sicherheit des Luftfahrzeuges nicht entscheidend. Die nötigen Abhilfemaßnahmen können von der Besatzung problemlos bewältigt werden. Geringfügige Ausfälle sind beispielsweise eine leichte Verringerung der Sicherheitstoleranzen oder der Funktionsfähigkeit, eine leichte Zunahme der Arbeitsbelastung für die Besatzung (z.B. Änderungen des Routineflugplans) sowie gewisse Unannehmlichkeiten für die Insassen.
E	<b>Ohne Wirkung:</b> Ein Ausfall hat keine Auswirkungen auf die Betriebsfähigkeit des Luftfahrzeuges und erhöht nicht die Arbeitsbelastung der Besatzung.

Tab. 38: Zertifizierungsstufen für die Folgen von Ausfällen im Flugverkehr [35]

Schmidtke beschreibt in seinem ergonomischen Bewertungsverfahren eine Einteilung nach dem potentiellen Schaden. Dazu dienen ihm drei Hauptstufen, welche wiederum in jeweils drei Unterstufen unterteilt sind [358]:

- I. Stufe 1-3: Bedingung ist unkritisch bzw. ohne wesentliche Einschränkungen voll zumutbar,
- II. Stufe 4-5: Bedingung ist teilweise kritisch bzw. nur unter Beachtung von Einschränkungen zumutbar,
- III. Stufe 6-9: Bedingung ist kritisch bzw. unzumutbar oder gesundheitsgefährdend.

Unkritisch bedeutet hierbei, dass sich „weder aus der Betrachtungsebene des Komforts noch aus der des Funktions- und Leistungszieles bzw. der Gesundheits- und Sicherheitsvorsorge nennenswerte Einschränkungen ergeben“. Von einer teilweise kritischen Bedingung geht Schmidtke aus, „wenn in der Komfortebene deutliche, in der Leistungs-, Gebrauchs- und Sicherheitsebene leichte bis deutli-

che, in der Gesundheitsebene noch keine Einschränkungen zu erwarten sind“ und von einer kritischen Bedingung, „wenn der Komfort schwer gefährdet und die Leistung, der Gebrauchswert, die Gesundheit und die Sicherheit gefährdet erscheint“. Schmidtke gibt an, dass für die Gewichtung der einzelnen Merkmale keine allgemein verbindlichen Empfehlungen vorzugeben sind. Beispielsweise sind die Geschwindigkeit und die Zugänglichkeit im Falle eines statischen Arbeitssystems nicht so wichtig wie beispielsweise bei einem Waffensystem, wo der Arbeitsplatz in kürzester Zeit einzunehmen und zu verlassen ist. Schmidtke beschreibt weiterhin für die Bewertungsstufen einige Bewertungsaussagen, welche neben dem kritischen Aspekt auch den Zumutbarkeitsgrad für den Menschen, den Mängelgrad für Arbeitsmittel, für Komfort-, Leistungs- und Gesundheitsbeeinträchtigung und ergonomische Empfehlungen beinhalten (Tab. 39).

Stufe	Allgemeine Bedienungs-kennzeichnung	Zumutbar-keitsgrad für den Men-schen	Mängelgrad für Arbeitsmittel	Komfort-, Leistungs- und Gesundheitsbeeinträch-tigung	Ergonomi-sche Empfehlung
1	Absolut unkri-tisch bzw. ausgezeichnet	Uneinge-schränkt zumutbar	Mängelfrei	Komfort, Leistung und Gesundheit nicht beeinträch-tigt	Uneinge-schränkt empfeh-lenswert
2	Unkritisch bzw. sehr gut	Voll zumut-bar	Mit sehr geringen Mängeln behaftet	Komfort leicht, Leistung und Gesundheit nicht beeinträch-tigt	Empfeh-lenswert
3	Überwiegend unkritisch bzw. gut	Zumutbar	Mit geringen Mängeln behaftet	Komfort deutlich, Leistung und Gesundheit nicht beeinträchtigt	Noch empfeh-lenswert
4	Im Allgemeinen noch unkritisch bzw. befriedi-gend	Mit leichten Einschränkungen zumutbar	Mit leichteren Mängeln behaftet	Komfort erheblich, Leistung leicht, Gesundheit nicht beeinträchtigt	Einge-schränkt empfeh-lenswert
5	Teilweise kritisch bzw. ausreichend	Einge-schränkt zumutbar	Mit erheblichen Mängeln behaftet	Komfort schwer, Leistung deutlich und Gesundheit nicht beeinträchtigt	Noch zulässig
6	Überwiegend kritisch bzw. unzureichend	Mit erheblichen Einschränkungen zumutbar	Mit bedenklichen Mängeln behaftet	Komfort schwerstens, Leistung erheblich und Gesundheit ggfs. leicht beeinträchtigt	Abzuraten
7	Kritisch bzw. mangelhaft	Weitgehend unzumutbar	Mit sehr bedenklichen Mängeln behaftet	Komfort schwerstens, Leistung schwer, Gesund-heit beeinträchtigt	Abzulehnen
8	Sehr kritisch bzw. ungenü-gend	Weitestge-hend unzumutbar	Mit schwerwie-genden Mängeln behaftet	Komfort, Leistung schwerstens und Gesund-heit beeinträchtigt	Unannehm-bar
9	In höchstem Maße kritisch bzw. völlig ungenügend	Absolut unzumutbar	Mit gefährlichen und/oder schwersten Mängeln behaftet	Komfort, Leistung und Gesundheit schwerstens gefährdet. Lebensgefahr	Unzulässig

**Tab. 39: Bewertungsstufen nach Schmidtke [358]**

Auch für den medizinischen Bereich gibt es verschiedene Klassifizierungsarten. Das BfArM klassifiziert eingehende Meldungen im Zusammenhang mit Medizinprodukten nach folgenden fünf Auswirkungsklassen [25]:

- Keine/vernachlässigbar: Befindlichkeitsstörungen, vorübergehende leichte Beschwerden,
- Gering: vorübergehende Verletzung oder Beeinträchtigung, die keine ärztliche / zahnärztliche Behandlung erfordert,
- . Schwerwiegend: Verletzung oder Beeinträchtigung, die eine ärztliche / zahnärztliche Behandlung erfordert, einschließlich semikritischer vorzeitiger Explantation,
- Kritisch: dauerhafte schwere Beeinträchtigung oder schwere Komplikation, die eine chirurgische oder intensivmedizinische Intervention erfordert, einschließlich kritischer vorzeitiger Explantation,
- Tod.

Das „National Coordinating Council for Medication Error Reporting and Prevention“ (NCC MERP) klassifiziert Probleme und Fehler nach dem entstandenen Schaden (Tab. 40).

<b>Kein Fehler</b>	<b>Kategorie A</b>	Umstände oder Fälle mit Fehlerpotential
<b>Fehler, kein Schaden<sup>57</sup></b>	<b>Kategorie B</b>	Ein Fehler tritt auf, erreicht jedoch nicht den Patienten. (Ein „Versäumnisfehler“ erreicht den Patienten.)
	<b>Kategorie C</b>	Ein Fehler erreicht den Patienten, verursacht jedoch keinen Schaden bei ihm.
	<b>Kategorie D</b>	Ein Fehler erreicht den Patienten, verlangt nach Beobachtung, um festzustellen, dass das Ergebnis kein Schaden am Patienten bedeutet, und/oder verlangt nach Intervention, um Schaden vorzubeugen.
<b>Fehler mit Schaden</b>	<b>Kategorie E</b>	Der aufgetretene Fehler ergibt einen temporären Schaden an einem Patienten oder trägt zu diesem bei und benötigt ein Eingreifen.
	<b>Kategorie F</b>	Der aufgetretene Fehler ergibt einen temporären Schaden an einem Patienten oder trägt zu diesem bei und erfordert einen anfänglichen oder anhaltenden Krankenhausaufenthalt.
	<b>Kategorie G</b>	Der aufgetretene Fehler ergibt einen permanenten Patientenschaden oder trägt zu diesem bei.
	<b>Kategorie H</b>	Der aufgetretene Fehler erfordert ein lebensrettendes Eingreifen.
<b>Fehler, Tod</b>	<b>Kategorie I</b>	Der aufgetretene Fehler trägt zum Tod des Patienten bei oder ist hierfür die Ursache.

Tab. 40: Fehlerkategorien nach dem entstandenen Schaden [adaptiert nach 281]

<sup>57</sup> Schaden ist in diesem Fall definiert als temporäre oder permanente Beeinträchtigung der physischen, emotionalen, psychischen Funktion oder des Körperbaus und/oder als daher resultierender Schmerz, welcher einen Eingriff verlangt [281].

Fehler werden in die vier Klassen eingeteilt: „kein Fehler“, „Fehler ohne Schaden“, „Fehler mit Schaden“ und „Tod“. Bei den ersten beiden Kategorien können jeweils noch Unterkategorien gewählt werden, sodass Fehler insgesamt in neun Kategorien unterteilt werden können. Bei der Wahl der Bewertungskategorie sollte der höchste Level des Schweregrades gewählt werden. Der Fall eines Patienten, der beispielsweise eine ernste anaphylaktische Reaktion erleidet (Kategorie H), eine Behandlung erhält (Kategorie F) und sich schließlich wieder erholt, sollte trotzdem mit der Kategorie H bewertet werden.

Die DIN EN IEC 60601-1-4 empfiehlt zur Bewertung des Schadensausmaßes für mögliche Konsequenzen einer Gefährdung die vier Kategorien: katastrophal (Möglichkeit von vielen Todesfällen oder schweren Verletzungen), kritisch (Möglichkeit eines Todesfalles oder einer schweren Verletzung), geringfügig (Möglichkeit einer Verletzung) und unwesentlich (geringe oder keine Möglichkeit einer Verletzung) [96]. Aber auch andere Einteilungen sind in Studien zu finden. Die israelische Forschergruppe Donchin et al. ließen für eine Datensammlung von menschlichen Fehlern bei medizinischem Personal mit Hilfe einer 5-Punkteskala Fehler hinsichtlich ihres Schweregrades bewerten. Die Skala gliederte sich von „least serious“ bis hin zu „most serious“. Hierbei wurden Fehler nach den schlimmstmöglichen Auswirkungen bewertet. Beispielsweise wurde eine Blutprobe zur Typisierung und zur Kreuzprobe mit einem falschen Patientenlabel versendet. Obwohl der Fehler bei der Blutbank sofort bemerkt wurde und keine Schädigung entstand, wurde er aufgrund der möglichen verhängnisvollen Auswirkungen als fatal und ernst eingestuft [119].

Cina et al. klassifizieren Medikationsfehler für Pharmazeutika nach dem Schweregrad in fünf Klassen: kein Schadenspotential, Schadenspotential, bedeutend (signifikant), gravierend und lebensbedrohlich [63].

## **6 Entwicklung des Bewertungsmaßstabes zur Bestimmung der Gebrauchstauglichkeit medizinischer Geräte**

Risikoabschätzungen für Gebrauchstauglichkeitsprobleme von Medizinprodukten sind bisher nur wenig betrachtet worden. Da die Gebrauchstauglichkeit von Medizinprodukten untrennbar mit Risiko verbunden ist, besteht die zwingende Notwendigkeit, auch dieses zu betrachten. Diese Forderung geht auch aus der DIN EN IEC 60601-1-6 (vgl. Kap. 2.3.2) hervor, die die gebrauchstauglichkeitsbedingte Sicherheit von medizinischen elektrischen Geräten sicherstellt. Zu diesem Zweck wird im Rahmen dieser Arbeit eine Bewertungsmöglichkeit für Gebrauchstauglichkeitsprobleme unter dem Aspekt der Sicherheit entwickelt.

Die Entwicklung dieser Bewertungsmethode erfolgte in einem iterativen Prozess, welcher von dem Prozess der Evaluationsmethode (vgl. Kap. 4) teilweise losgelöst war. Zuerst wurden auf Basis der in Kap. 5 beschriebenen Grundlagen vier verschiedene Konzepte erarbeitet. Ein interdisziplinäres Expertenteam prüfte daraufhin die Relevanz der einzelnen Methoden für den medizinischen Bereich, um die weiterführende Methode auszuwählen. Diese wurde, wie die Standardprozedur, in einem iterativen Entwicklungsprozess weiterentwickelt und konkretisiert. Potentielle Anwender der Methode und Fachexperten wurden in die Entwicklung mit einbezogen. Im Folgenden werden die einzelnen Konzepte vorgestellt.

### **6.1 Verschiedene Bewertungskonzepte**

#### **6.1.1 Konzept 1 – 3D-Matrix**

Das erste Konzept orientiert sich an der in Kapitel 5.4.2 beschriebenen Risikomatrix. Gebrauchstauglichkeitsprobleme werden, wie in Kapitel 5.8.2 beschrieben, durch den Schweregrad des Problems und seine Auftretenshäufigkeit definiert [72, 124, 184, 284, 289, 348]. Risiko wird durch eine Kombination aus der Wahrscheinlichkeit, mit der ein Schaden auftritt, und dem Ausmaß dieses Schadens definiert [86, 100, 103, 108, 234]. Das erste Konzept stützt sich auf eine Kombination dieser beiden Ansätze (Abb. 84). Durch eine Kombination der drei Achsen ergibt sich ein dreidimensionales Konstrukt (Matrix), womit Gebrauchstauglichkeitsprobleme im medizinischen Kontext bewertet werden können. Im Folgenden wird auf die Bestimmung der einzelnen Faktoren genauer eingegangen:





**Abb. 84: Kombination der Gebrauchstauglichkeitsbewertung und der Risikobewertung für ein Konzept zur Bewertung der Gebrauchstauglichkeit medizinischer Produkte**

### Auftretenswahrscheinlichkeit

Die Auftretenswahrscheinlichkeit eines Problems wird sowohl bei der Risikobewertung als auch bei der Bewertung von Gebrauchstauglichkeitsproblemen betrachtet. In der Regel wird bei quantitativen Einstufungen in beiden Bereichen eine Einteilung mit Schlagworten (Deskriptoren) verwendet [100, 102, 124, 284, 348]. Für die Einstufung der Auftretenshäufigkeit von Gebrauchstauglichkeitsproblemen im medizinischen Kontext wurde im Rahmen dieses Konzeptes eine siebenstufige Skala erstellt (Tab. 41).

Auftreten	Beschreibung	Beleg aus Nutzertest		
		Primäraufgaben	Sekundäraufgaben (häufig verwendet)	Sekundäraufgaben (selten verwendet)
Unglaublich	Es ist in hohem Maße unwahrscheinlich, dass das Problem auftritt.	Das Problem wurde bei Expertenverfahren entdeckt, trat aber im Gebrauchstauglichkeitstest niemals auf und wurde von Experten als höchst unwahrscheinlich eingestuft.		
Unwahrscheinlich	Die Wahrscheinlichkeit, dass das Problem auftritt, ist extrem gering, jedoch ist es unter speziellen Umständen möglich.	Das Problem wurde bei Expertenverfahren entdeckt, trat aber im Gebrauchstauglichkeitstest niemals auf und wurde von Experten als unwahrscheinlich eingestuft.		
Entfernt	Das Problem wird wahrscheinlich auftreten.	Das Problem wurde bei Expertenverfahren entdeckt, trat aber im Gebrauchstauglichkeitstest niemals auf, wurde dennoch von Experten als relevant betrachtet.	Ein Testnutzer wurde mit dem Problem konfrontiert (oder < 15% der Testnutzer).	
Gelegentlich	Das Problem wird wahrscheinlich einige Male auftreten.	< 15% der Testnutzer werden mit dem Problem konfrontiert.	< 30% der Testnutzer werden mit dem Problem konfrontiert.	< 40% der Testnutzer werden mit dem Problem konfrontiert.
Wahrscheinlich	Das Problem wird wahrscheinlich bei einigen Nutzern auftreten.	≥ 15% der Testnutzer werden mit dem Problem konfrontiert.	≥ 30% der Testnutzer werden mit dem Problem konfrontiert.	≥ 40% der Testnutzer werden mit dem Problem konfrontiert.
Häufig	Das Problem wird öfters auftreten. Gerade unerfahrene Benutzer werden häufig mit dem Problem konfrontiert.	≥ 40% der Testnutzer werden mit dem Problem konfrontiert.	≥ 50% der Testnutzer werden mit dem Problem konfrontiert.	≥ 80% der Testnutzer werden mit dem Problem konfrontiert.
Regelfall	In der Regel wird das Problem auftreten.	≥ 60% der Testpersonen werden mit dem Problem konfrontiert.	≥ 80% der Testnutzer werden mit dem Problem konfrontiert.	Nicht möglich

**Tab. 41: Einstufung der Auftretenswahrscheinlichkeit eines Gebrauchstauglichkeitsproblems**

Die qualitativen Abstufungen wurden zusätzlich durch Beschreibungen und quantitative Belege aus der Gebrauchstauglichkeitsevaluation ergänzt. Hierbei wurde die in Kapitel 1.2.2.2 verwendete Einteilung der Aufgaben in Primär- und Sekundäraufgaben berücksichtigt. Es ist davon auszugehen, dass eine Primäraufgabe mit höherer Wahrscheinlichkeit von einem Benutzer verwendet wird als eine Sekundäraufgabe.

### Schweregrad des Gebrauchstauglichkeitsmerkmals

Die Bewertung des Problemschweregrades orientiert sich in diesem Konzept an den in der Praxis meist verwendeten 3-4teiligen Einstufungen, um Gebrauchstauglichkeitsprobleme nach der Priorität zu sortieren [124, 284, 348]. Um dem Charakter einer Bewertung jedoch gerecht zu werden, müssen auch besonders positive Beispiele berücksichtigt werden, die sich aus der Masse herausheben. Hierzu wurde die an Nielsen angelehnte [284] sechsstufige Skala durch die Abstufung „hervorragende Lösung“ ergänzt (Tab. 42).

Aufgabeneinstufung	Beschreibung	Maße aus Nutzertest
Hervorragende Lösung	Die Aufgabe wird in einer überdurchschnittlich geringen Zeit erfüllt. Der Nutzer ist mit der Aufgabenerfüllung sehr zufrieden.	Hohe Effektivität, Effizienz und Zufriedenheit
Kein Problem	Es fällt kein Problem auf. Die Aufgabe wird in einer durchschnittlich geringen Zeit erfüllt. Der Nutzer ist mit der Aufgabenerfüllung zufrieden.	Effektivität, Effizienz und Zufriedenheit werden erfüllt.
Kosmetisches Problem	Probleme, die leicht störend wirken, aber die Gebrauchstauglichkeit, wenn auch nur gering, beeinflussen.	Effektivität erfüllt, geringfügige Abstriche in Effizienz und/oder Zufriedenheit sind feststellbar.
Geringes Problem	Probleme haben geringfügigen Einfluss auf die Gebrauchstauglichkeit. Der Nutzer wird ggf. das Problem schnell bemerken und kann es schnell beheben.	Effektivität erfüllt, Abstriche in Effizienz und Zufriedenheit.
Mittleres Problem	Problem erzeugt Verzögerung und Frustration.	Effektivität erfüllt bzw. Aufgabe kann durch Workaround <sup>58</sup> fortgesetzt werden, verringerte Effizienz und Zufriedenheit.
Großes Problem	Problem erzeugt signifikante Verzögerung und Frustration.	Effektivität erfüllt bzw. Aufgabe kann durch Workaround fortgesetzt werden. Schlechte Effizienz und Zufriedenheit.
Katastrophe	Problem hindert an der Aufgabenerfüllung.	Effektivität nicht erfüllt, Workaround nicht möglich bzw. wird vom Nutzer nicht gefunden.

Tab. 42: Einstufung einer Arbeitsaufgabe und eines Gebrauchstauglichkeitsproblems

<sup>58</sup> Unter einem Workaround versteht man eine Umgehungslösung (Notbehelf). Ein erkanntes Problem wird durch eine Hilfskonstruktion umgangen, die das Problem aber nicht löst.

### Klassifizierung nach potentiellm Schaden

Da diese Methode die Sicherheit von Geräten präventiv betrachtet, können auch bei der Abschätzung eines potentiellen Schadens lediglich qualitative Einstufungen verwendet werden. In der Risikobetrachtung werden verschiedene, mehrstufige Kategorisierungsmethoden vorgeschlagen (vgl. Kap. 5.4.2). Für die Betrachtung von Schadensauswirkungen in der Medizin ist es jedoch sinnvoll, ein Kategorisierungssystem aus dem medizinischen Fachbereich zu verwenden (vgl. Kap. 5.1.5). Um die, für die Risikobetrachtung gewünschte geringe Komplexität der Methode nicht durch zu viele Abstufungen, wie z.B. bei der 9-stufigen Kategorisierung des NCC MERP, zu gefährden (vgl. Kap. 5.8.5), wurde eine 4-stufige Kategorisierung verwendet. Deren Wortlaut ist an Kategorisierungsschemata aus dem medizinischen Bereich angelehnt [281]. Sie genügt den Anforderungen an Schemata des Risikomanagements (Tab. 43).

<b>Stufe I</b>	Keine Auswirkung auf die Gesundheit des Patienten und des Personals bzw. keine Schäden am Material
<b>Stufe II</b>	Mögliche gefährliche Situation, die Sachschäden oder leichte Körperverletzungen zur Folge haben kann, wenn sie nicht vermieden wird
<b>Stufe III</b>	Mögliche gefährliche Situation, die mittlere Körperverletzungen zur Folge haben kann, wenn sie nicht vermieden wird
<b>Stufe IV</b>	Mögliche gefährliche Situation, die schwere Körperverletzungen oder Tod zur Folge haben kann, wenn sie nicht vermieden wird

**Tab. 43: 4-stufige Kategorisierung nach dem potentiellen Schaden**

Zur Verknüpfung der drei Einflussgrößen für Gebrauchstauglichkeitsprobleme im medizinischen Kontext wurde eine 3-dimensionale Matrix gewählt, welche die einzelnen Kombinationen zu einer 10-stufigen Bewertung zusammenführt (Abb. 85). Hierdurch entstehen theoretisch 196 Kombinationen ( $7 \times 7 \times 4$ ). Es sind jedoch nur 133 praktisch möglich, da bei der Einstufung einer Aufgabe mit „Hervorragende Lösung“, „Kein Problem“, oder „Kosmetisches Problem“ keine Gefährdung eines Patienten durch ein Gebrauchstauglichkeitsproblem vorkommen kann (Tab. 44).

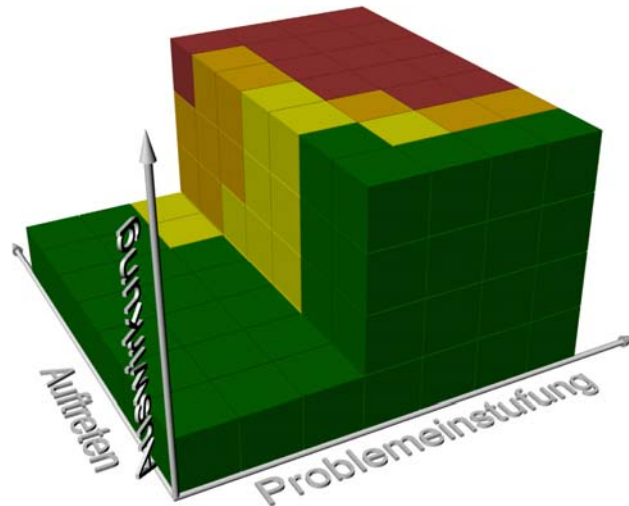


Abb. 85: 3-dimensionale Matrix zur Bewertung der Gebrauchstauglichkeit medizinischer Produkte

Keine Auswirkung auf die Gesundheit des Patienten							
Einstufung \ Auftreten	Hervorra-gende Lösung	Kein Usability Problem	Kosmetisches Problem	Geringes Usability Problem	Mittleres Usability Problem	Großes Usability Problem	Usability Katastrophe
Un glaublich	X	IX	IX	IX	IX	IX	VIII
Un wahr-scheinlich	X	IX	IX	IX	VIII	VIII	VII
Entfernt	X	IX	IX	VIII	VII	VII	VI
Gelegentlich	X	IX	IX	VII	VII	VI	V
Wahrscheinlich	X	IX	VIII	VII	VI	VI	V
Häufig	X	IX	VII	VII	VI	V	IV
Regelfall	X	IX	VII	VII	VI	V	IV
Mögliche gefährliche Situation, die Sachschäden oder leichte Körperverletzungen zur Folge haben kann							
Un glaublich	n/a	n/a	n/a	IX	IX	VIII	VIII
Un wahr-scheinlich	n/a	n/a	n/a	VIII	VI	VI	V
Entfernt	n/a	n/a	n/a	VII	V	IV	IV
Gelegentlich	n/a	n/a	n/a	VII	IV	III	III
Wahrscheinlich	n/a	n/a	n/a	VI	IV	III	III
Häufig	n/a	n/a	n/a	V	III	III	III
Regelfall	n/a	n/a	n/a	V	III	III	III
Mögliche gefährliche Situation, die mittlere Körperverletzungen zur Folge haben kann							
Un glaublich	n/a	n/a	n/a	IX	IX	VIII	VIII
Un wahr-scheinlich	n/a	n/a	n/a	VIII	VI	VI	V
Entfernt	n/a	n/a	n/a	VII	V	IV	III
Gelegentlich	n/a	n/a	n/a	VI	IV	II	II
Wahrscheinlich	n/a	n/a	n/a	V	III	II	II
Häufig	n/a	n/a	n/a	V	II	II	II
Regelfall	n/a	n/a	n/a	V	II	II	II
Mögliche gefährliche Situation, die schwere Körperverletzungen oder Tod zur Folge haben kann							
Un glaublich	n/a	n/a	n/a	IX	IX	VIII	VII
Un wahr-scheinlich	n/a	n/a	n/a	VII	VI	V	IV
Entfernt	n/a	n/a	n/a	VI	V	III	III
Gelegentlich	n/a	n/a	n/a	VI	III	II	II
Wahrscheinlich	n/a	n/a	n/a	V	II	I	I
Häufig	n/a	n/a	n/a	IV	I	I	I
Regelfall	n/a	n/a	n/a	III	I	I	I

Tab. 44: Bewertungskonzept der 3-D Matrix (n/a = nicht anwendbar).

Die zehn Bewertungsstufen wurden anhand der entsprechenden Maßstäbe für Gebrauchstauglichkeit formuliert und mit dem Aspekt der Sicherheit ergänzt (Tab. 45). Zusätzlich wurden die Bewertungsstufen farblich kodiert, sodass eine vierstufige farbige Unterteilung entstand, welche an ein bekanntes Ampelschema angelehnt ist. Die farbigen Bereiche grün bis orange spiegeln hierbei den reinen Gebrauchstauglichkeitsbereich wider, d.h. den Bereich, bei dem eine durch Gebrauchstauglichkeitsprobleme verursachte Gefährdung unwahrscheinlich ist. Im Sinne der DIN EN IEC 60601-1-6 und der DIN EN IEC 62366 können Geräte, welche Probleme in diesem Bereich aufweisen, als normkonform angesehen werden. Im roten Bereich können die Gebrauchstauglichkeitsprobleme einen Einfluss auf die Sicherheit haben. Würde ein Problem in diesen Bereich eingeordnet werden, so würde man von einer Nichtkonformität mit diesen Normen sprechen.

Normkonform	X	Medizinisches Gerät ist herausragend, Bedienbarkeit und Benutzerfreundlichkeit überraschen positiv.
	IX	Medizinisches Gerät ist gut und erzeugt Zufriedenheit. Die Erwartungen bezüglich Bedienbarkeit und Benutzerfreundlichkeit werden weitestgehend erfüllt.
	VIII	Medizinisches Gerät ist gut. Gebrauchstauglichkeitsprobleme sind sehr gering und treten selten auf. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich.
	VII	Medizinisches Gerät ist befriedigend, irritiert aber in einzelnen Fällen. Geringe Gebrauchstauglichkeitsprobleme sind festzustellen. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich.
	VI	Medizinisches Gerät erfüllt die Erwartungen nicht vollständig, es ist nur ausreichend. Gebrauchstauglichkeitsprobleme sind feststellbar. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich.
	V	Medizinisches Gerät erzeugt Unzufriedenheit. Gebrauchstauglichkeitsprobleme treten häufig auf. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind unwahrscheinlich.
	IV	Medizinisches Gerät ist mangelhaft und führt zur Enttäuschung. Gebrauchstauglichkeitsprobleme treten häufig auf. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind unwahrscheinlich.
Nicht normkonform	III	Medizinisches Gerät verunsichert bzw. verursacht extreme Verärgerung. Gebrauchstauglichkeitsprobleme treten häufig auf und können zu gefährlichen Situationen führen. Sachschäden oder Körperverletzungen können folgen, wenn auftretende Probleme nicht erkannt und behoben werden.
	II	Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsmängel. Dadurch verursachte gefährliche Situationen mit Sach- oder Personenschäden sind nicht auszuschließen, sollten die auftretenden Probleme nicht erkannt und behoben werden.
	I	Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsmängel. Es stellt ein Sicherheitsrisiko für den Patienten und/oder die Umgebung dar, sollten die auftretenden Probleme nicht erkannt und behoben werden.

Tab. 45: Bewertungsstufen für die Gebrauchstauglichkeit medizinischer Geräte

### 6.1.2 Konzept 2 - Klassifizierung

Das zweite Konzept ist durch die von Denoix entwickelte TNM-Klassifizierung zur Tumorstadieneinteilung inspiriert [419]. Wie im ersten Konzept wurden als bewertende Faktoren die Häufigkeit und die Folgen von Problemen betrachtet. Unterschiedlich ist, dass hierbei zwischen medizinischen und wirtschaftlichen Folgen differenziert wurde. Zusätzlich wurde noch der Faktor Zufriedenheit eingeführt, da dieser in den wirtschaftlichen Folgen, welche durch Effizienz und Effektivität abgedeckt sind, nicht repräsentiert war. Als Maß für Effektivität und Effizienz wurde nur der zeitliche Aspekt berücksichtigt. Die einzelnen Faktoren wurden beschrieben (Tab. 46) und im späteren Verlauf mit einem Faktor gewichtet (Tab. 47), um schließlich mit Hilfe der Formel:

$$(F_{Wi} + Z_i) \times H_i + H_i \times F_{Mi} = H_i \times (F_{Wi} + Z_i + F_{Mi})$$

eine Bewertung zu ermöglichen (Tab. 48).

<b>H = Häufigkeit</b>		<b>Z = Zufriedenheit</b>	
H <sub>0</sub> =	Es tritt kein Problem auf	Z <sub>0</sub> =	zufrieden
H <sub>1</sub> =	Problemhäufigkeit ≤5%	Z <sub>1</sub> =	verstört
H <sub>2</sub> =	Problemhäufigkeit 6-15%	Z <sub>2</sub> =	genervt
H <sub>3</sub> =	Problemhäufigkeit 16-50%	Z <sub>3</sub> =	verärgert
H <sub>4</sub> =	Problemhäufigkeit 51-80%		
H <sub>5</sub> =	Problemhäufigkeit 80-100%		
<b>F = Folgen</b>			
<u>F<sub>M</sub> = Medizinische Folgen</u>			
F <sub>M0</sub> =	keine medizinischen Folgen		
F <sub>M1</sub> =	temporärer Schaden am Patienten oder Personal		
F <sub>M2</sub> =	temporärer Schaden am Patienten oder Personal, welcher einen anfänglichen oder anhaltenden Krankenhausaufenthalt erfordert		
F <sub>M3</sub> =	permanenten Schaden am Patienten oder Personal		
F <sub>M4</sub> =	Schaden am Patienten oder Personal erfordert lebensrettendes Eingreifen.		
F <sub>M5</sub> =	Schaden trägt zum Tod des Patienten oder des Personals bei oder ist die Ursache für diesen.		
<u>F<sub>W</sub> = Wirtschaftliche Folgen (Effektivität und Effizienz)</u>			
F <sub>W0</sub> =	keine wirtschaftlichen Folgen		
F <sub>W1</sub> =	Aufgabe dauert 10-50% länger als der Normalfall <sup>59</sup>		
F <sub>W2</sub> =	Aufgabe dauert 50-100% länger als der Normalfall		
F <sub>W3</sub> =	Aufgabe dauert 2-3 mal länger als der Normalfall		
F <sub>W4</sub> =	Aufgabe dauert mehr als 3 mal länger als der Normalfall		
F <sub>W5</sub> =	Revisions-OP notwendig		

**Tab. 46: Faktoren für die Klassifizierung von Gebrauchstauglichkeitsproblemen für das 2. Konzept**

<sup>59</sup> Normalfall: Wie lange benötigt der Produktmanager? Wie lange dauert die Aufgabe ohne Probleme?

**Punktesystem**

Häufigkeit	Faktor	Folgen (wirtschaftlich)	Faktor	Folgen (medizinisch)	Faktor	Zufriedenheit	Faktor
H <sub>0</sub>	<b>0</b>	F <sub>W0</sub>	<b>0</b>	F <sub>M0</sub>	<b>1</b>	Z <sub>0</sub>	<b>0</b>
H <sub>1</sub>	<b>0,5</b>	F <sub>W1</sub>	<b>1</b>	F <sub>M1</sub>	<b>10</b>	Z <sub>1</sub>	<b>1</b>
H <sub>2</sub>	<b>0,75</b>	F <sub>W2</sub>	<b>2</b>	F <sub>M2</sub>	<b>100</b>	Z <sub>2</sub>	<b>2</b>
H <sub>3</sub>	<b>1</b>	F <sub>W3</sub>	<b>4</b>	F <sub>M3</sub>	<b>200</b>	Z <sub>3</sub>	<b>3</b>
H <sub>4</sub>	<b>2</b>	F <sub>W4</sub>	<b>8</b>	F <sub>M4</sub>	<b>400</b>		
H <sub>5</sub>	<b>4</b>	F <sub>W5</sub>	<b>10</b>	F <sub>M5</sub>	<b>800</b>		

**Tab. 47: Gewichtungsfaktoren**

Punkte	
0	Medizinisches Gerät ist herausragend, Bedienbarkeit und Benutzerfreundlichkeit sind ausgezeichnet.
1-2	Medizinisches Gerät ist gut, geringe Gebrauchstauglichkeitsprobleme sind feststellbar.
3-5	Medizinisches Gerät weist Gebrauchstauglichkeitsprobleme auf.
5-10	Große Gebrauchstauglichkeitsprobleme sind feststellbar.
10-31	Gebrauchstauglichkeits-Katastrophe
32-131	Gebrauchstauglichkeitsprobleme mit geringem Risiko für Personal und Patienten sind feststellbar.
132-231	Gebrauchstauglichkeitsprobleme mit Risiko für Personal und Patienten sind feststellbar.
232-400	Gebrauchstauglichkeitsprobleme mit hohem Risiko für Personal und Patienten sind feststellbar.
401-600	Gebrauchstauglichkeitsprobleme mit nicht mehr tolerierbaren Risiken für Personal oder Patienten sind feststellbar.
>600	Gebrauchstauglichkeitsprobleme mit nicht mehr tolerierbaren Risiken für Personal oder Patienten, bis hin zum Tod, sind feststellbar.

**Tab. 48: Ergebnisstufen für das 2. Konzept****6.1.3 Konzept 3 - Nutzwertanalyse**

Das dritte Konzept bedient sich der Überlegungen aus der Nutzwertanalyse. Nach Zangmeier ist eine Nutzwertanalyse (NWA) die „Analyse einer Menge komplexer Handlungsalternativen mit dem Zweck, die Elemente dieser Menge entsprechend den Präferenzen des Entscheidungsträgers bezüglich eines multidimensionalen Zielsystems zu ordnen. Die Abbildung der Ordnung erfolgt durch die Angabe der Nutzwerte (Gesamtwerte) der Alternativen" [424]. Sie dient dazu, unter schwer differenzierbaren Alternativen auszuwählen und ist besonders geeignet, wenn „weiche“, also quantitativ nicht erfassbare Größen vorliegen. Durch die Abbildung parametrisierter Einflussgrößen kann eine Entscheidung getroffen werden. Die Nutzwertanalyse beachtet, dass Einflussgrößen eine unterschiedliche Relevanz haben. Damit diese vergleichbar werden, müssen sie zu dem sog. Erfül-

lungsgrad vereinheitlicht werden. Um die beeinflussenden Kriterien zu bewerten, werden Punktwerte von 0 bis 4 vergeben.

0 → Kriterium voll erfüllt

1 .

2 .

3 .

4 → Kriterium nicht erfüllt

Danach wird den einzelnen Kriterien eine Bedeutung zugewiesen, welche durch Gewichtungsfaktoren ausgedrückt wird. Die Gewichte werden hierbei so gewählt, dass ihre Summe 100% ergibt. Für die Gewichtung der einzelnen Kriterien ist bei diesem Konzept der Anwender selbst verantwortlich. Er verteilt Gewichtspunkte von 0 „irrelevant“ bis 10 „extrem hohe Bedeutung“. Zur Normierung der einzelnen Gewichtungsfaktoren auf den Wert 100% werden die einzelnen Kriteriengewichte durch die Summe aller Kriteriengewichte dividiert. Als beeinflussende Faktoren werden:

- geringe Auftretenshäufigkeit des Problems
- kein Schaden durch das Problem
- Zufriedenheit der Nutzer
- Aufgabenerfüllung

als Minimalfaktoren vorgegeben, welche jedoch geräteabhängig ergänzt werden können. Im Folgenden wird dieses Konzept anhand eines Beispiels veranschaulicht:

#### 1. Bestimmung der beeinflussenden Kriterien:

- geringe Auftretenshäufigkeit des Problems
- kein Schaden durch das Problem
- Zufriedenheit der Nutzer
- Aufgabenerfüllung

#### 2. Gewichtung und Normierung der Kriterien

Als wichtigstes Kriterium wird der Schaden, der durch das Problem verursacht wird, gesehen. Daher wird dieses Kriterium mit dem höchsten Faktor „10“ gewichtet. Die Aufgabenerfüllung und die Auftretenshäufigkeit des Problems haben eine mittlere Gewichtung (Faktor „5“). Als relativ unwichtig wird die Zufriedenheit des Nutzers gesehen, daher Faktor „2“.

Kriterium	Faktor	Normierter Faktor
Kein Schaden durch das Problem	10	0,45 (10/22)
Aufgabenerfüllung	5	0,23 (5/22)
Geringe Auftretenshäufigkeit des Problems	5	0,23 (5/22)
Zufriedenheit der Nutzer	2	0,09 (2/22)
	∑ 22	∑ 1



**3. Bewertung der Kriterien**

Das Problem erreichte zwar den Patienten, verursachte allerdings nur eine leichte Körperverletzung, die noch reparabel war. Daher wird das Kriterium mit 1 bewertet. Die Aufgabe dauerte jedoch um das Doppelte länger als im Normalfall und verursachte eine Verärgerung bei den Benutzern. Daher werden diese beiden Kriterien mit „3“ bewertet. Das Problem trat bei einer Aufgabe auf, die selten ausgeführt wird, daher wird die Auftretenshäufigkeit mit „2“ beziffert.

Kriterium	Bewertung
Kein Schaden durch das Problem	1
Aufgabenerfüllung	3
Geringe Auftretenshäufigkeit des Problems	2
Zufriedenheit der Nutzer	3

0	→ Kriterium voll erfüllt
1	.
2	.
3	.
4	→ Kriterium nicht erfüllt

**4. Berechnung des Problemwertes**

Um den Problemwert zu erhalten, wird die Summe der einzelnen Kriterienprodukte berechnet:

$$\sum (\text{Kriterienbewertung} \times \text{Gewichtung})$$

Beispiel:

Kriterium	Gewicht	Bewertung	Zielertrag
Kein Schaden durch das Problem	0,45	1	0,45
Aufgabenerfüllung	0,23	3	0,69
Geringe Auftretenshäufigkeit des Problems	0,23	2	0,46
Zufriedenheit der Nutzer	0,09	3	0,27
			$\sum$ 1,87

Die Ergebnisse dieser Methode liegen in einem Wertebereich zwischen 0 und 4. Je höher der Wert ist, desto wichtiger wird das Problem. Die Ergebnisse dieser Methode können, wie auch bei den anderen Konzepten, in eine Ergebnistabelle einsortiert werden.

**6.1.4 Konzept 4 – Risikograph**

Das vierte Konzept orientiert sich an dem in Kapitel 5.4.2 beschriebenen Risikographen. Im Gegensatz zu der im Konzept 1 verwendeten Risikomatrix wird hier, zusätzlich zu den Faktoren „Auftretenswahrscheinlichkeit“, „Problemeinstufung“ und „Schadensausmaß“, der Faktor „Gefahrenabwendung“ mit den Ausprägungen „nicht möglich“ und „möglich unter bestimmten Bedingungen“ eingeführt (Tab. 49). Das Ergebnis ist eine Einteilung in verschiedene Bewertungsklassen, welche beispielhaft in Abb. 86 dargestellt ist. In den Zweigen S2 und S3, also den Zweigen, in denen ein Schaden den Patienten erreichen kann, wird die Ausprägung der Problemeinstufung „kosmetisches Problem“ herausgenommen, da bei einem derartigen Problem kein Schaden beim Patienten zu erwarten ist. In dem beschriebenen Konzept entsprechen die Ergebnisse den in Konzept 1 vorgeschlagenen Bewertungen.

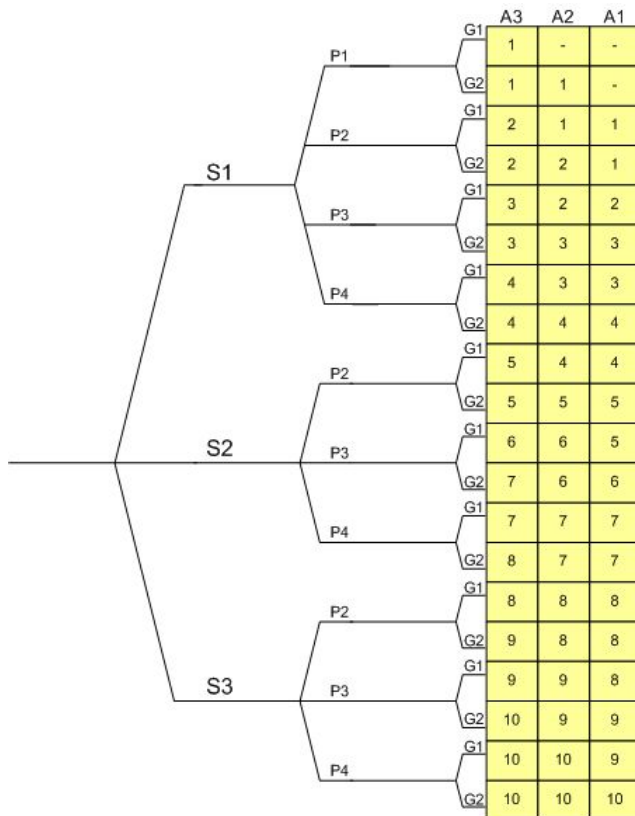


Abb. 86: Risikograph zur Bewertung der Gebrauchstauglichkeit medizinischer Produkte

Risikoparameter	Ausprägungen	
Problemeinstufung	P1	kosmetisches Problem
	P2	geringes Problem
	P3	mittleres Problem
	P4	großes Problem
Auftrittenswahrscheinlichkeit	A1	sehr gering
	A2	gering
	A3	hoch
Schadensausmaß	S1	kein Schaden
	S2	leichte Verletzung einer Person
	S3	schwere irreversible Verletzungen bis hin zum Tod
Gefahrenabwendung	G1	möglich unter bestimmten Bedingungen
	G2	nicht möglich

Tab. 49: Risikoparameter des Risikographens

## 6.2 Auswahl, Entwicklung und Evaluation der Bewertungsmethode

Um ein allgemeingültiges Bewertungssystem für die Gebrauchstauglichkeit in der Medizin zu entwickeln, wurde analog zur Entwicklung der Standardprozedur ein iterativer Entwicklungsprozess durchgeführt. In der ersten Phase wurde die weiterzuverfolgende Methode anhand der in Kapitel 3.3 aufgestellten Anforderungen überprüft, um sie in anschließenden Studien und Evaluationen weiterzuentwickeln. Hinzu kommen noch Anforderungen, welche aus den in diesem Kapitel beschriebenen Erfahrungen entstanden sind und nur für die Bewertung gelten. Diese ergänzen die 15 bereits aufgestellten Anforderungen:

### A16. Bewertung nach Normkonformität mit der DIN EN IEC 60601-1-6 und DIN EN IEC 62366

Die Methode soll Ergebnisse liefern, die Rückschlüsse auf eine Normkonformität mit den Sicherheitskriterien der Gebrauchstauglichkeitsnormen für Medizingeräte DIN EN IEC 60601-1-6 und DIN EN 62366 erlauben.

#### **A17. Allgemeinverständliches Bewertungssystem**

Die Methode soll ein Bewertungsschema liefern, welches schnell und intuitiv verstanden wird.

#### **A18. Verknüpfung von Gebrauchstauglichkeits- und Risikobewertung**

Die Methode soll eine Gebrauchstauglichkeitsbewertung unter Berücksichtigung des sicherheitskritischen Aspekts durchführen.

#### **A19. Bezug zu den entsprechenden Anforderungen der relevanten Normen für Sicherheit und Gebrauchstauglichkeit**

Die Methode soll bzgl. Gebrauchstauglichkeit sowohl den Anforderungen der Standardnormen [105, 113, 114, 200] als auch den Anforderungen der speziell auf medizinische Geräte ausgerichteten Normen [97, 104] und denen des Risikomanagements in der Medizin [108] entsprechen.

#### **A20. Aufbau der Bewertung auf den Ergebnissen der Standardprozedur**

Die Bewertungsmethode soll die Ergebnisse der in dieser Arbeit entwickelten Standardprozedur verwenden, um die Gebrauchstauglichkeit eines medizinischen Gerätes zu beurteilen.

### **6.2.1 Entwicklungsstudie 1 – Auswahl der Methode**

Der entwickelte Anforderungskatalog gilt als Bewertungsgrundlage für die Konzeptauswahl und die Validierung der Methode. Der Anforderungskatalog umfasst sowohl Anforderungen an Ziele, welche mit dem Bewertungsmaßstab erreicht werden sollen, als auch Anforderungen an die Verwendung des Bewertungsmaßstabes und der Evaluationsmethode. In der ersten Entwicklungsstudie wurden in einem Gespräch mit einem Experten auf dem Gebiet der Gebrauchstauglichkeit in der Medizin, die Konzepte auf ihre Relevanz für die in dieser Arbeit behandelten Fragestellungen kritisch begutachtet. Daraufhin wurde das weiter zu verfolgende Konzept ausgewählt.

Das **erste Konzept** ist direkt aus der Gebrauchstauglichkeits- und der Sicherheitsbetrachtung abgeleitet. Es verbindet die jeweiligen Einflussfaktoren dieser Bereiche. Die verwendete Matrixmethode hat sich in einer zweidimensionalen Form bereits bei der Sicherheitsbetrachtung in vielen Bereichen bewährt. Dennoch umfasst diese Methode nicht die komplette Problematik im Zusammenhang mit Gebrauchstauglichkeit, weshalb für dieses Konzept noch eine dritte Dimensi-

on eingeführt wurde. Es bleibt zu überprüfen, ob das Konzept hierdurch nicht zu komplex im Sinne der Anwenderfreundlichkeit ist.

Das **zweite Konzept** lässt keine genaue Betrachtung der für die Gebrauchstauglichkeit wichtigen Maße Effektivität und Effizienz zu. Bereits im konzeptionellen Stadium wirkte es kompliziert, ein Negativaspekt der sich durch eine Integration der Gebrauchstauglichkeitsmaße noch zusätzlich verstärken würde. Da diese Methode von interdisziplinären Experten durchgeführt werden soll, kann nicht davon ausgegangen werden, dass das entsprechende Vorwissen für diese Methode generell vorhanden ist.

Das **dritte Konzept** ähnelt dem ersten, jedoch führt hier der Bewerter selbst eine Gewichtung mit Hilfe nicht standardisierter Faktoren durch. Dieser hohe Freiheitsgrad ist für die Reliabilität nicht förderlich. Mit vorgegebenen Maßen, Definitionen und Einstufungen wird nämlich sichergestellt, dass Anwender unter den unterschiedlichen Kriterien dasselbe verstehen.

Das **vierte Konzept** verbindet analog zum ersten Konzept Sicherheit und Gebrauchstauglichkeit. Es wird angewendet, wenn zu viele Faktoren die zu betrachtende Problematik beeinflussen. Die Methode des Risikographen hat sich in vielen Bereichen zur Abschätzung von Risiken bewährt, jedoch wächst auch hier mit zunehmender Faktorenvielesart die Komplexität der Methode.

Als geeignet wurden die Konzepte eins und vier empfunden, da sie das höchste Potential boten, um die aufgestellten Anforderungen zu erfüllen. Konzept eins erhielt hierbei den Vorzug für die Weiterentwicklung, da es bei drei beeinflussenden Faktoren (Aufretenshäufigkeit, Problemschwere und Schadensausmaß) verständlicher wirkte. Konzept vier wäre geeignet, wenn mehr als drei Faktoren die Gebrauchstauglichkeit medizinischer Produkte beeinflussen würden. In den folgenden Evaluationen wird überprüft, ob sich diese Einschätzung bewahrheitet. Für die Weiterentwicklung wird die Methode sehr allgemein gehalten, da sie in ihrer Grundform für alle medizinischen Geräte gelten soll.

## **6.2.2 Entwicklungsstudie 2 - Evaluation der Bewertungsmethode**

### **Fragestellung:**

In der zweiten Entwicklungsstudie wurde die Methode erstmals praktisch eingesetzt, um sie zu evaluieren und das Verbesserungspotential ableiten zu können. Hierbei sollten aufgedeckte Gebrauchstauglichkeitsprobleme besonders unter dem Aspekt der Sicherheit bewertet werden, um im Anschluss eine Bewertung

der Gebrauchstauglichkeit des gesamten Produktes durchführen zu können. Die Methode sollte ihre einfache Anwendbarkeit unter Beweis stellen und ein sinnvolles und allgemein verständliches Ergebnis liefern. Auch sollte die Konformität der Methode mit der Zusatznorm DIN EN IEC 60601-1-6 festgestellt werden.

Ziel war es auch herauszufinden, welcher Personenkreis zur Anwendung der Methode geeignet ist. Hierzu wurden die Interessensvertreter ermittelt und in Gruppen unterteilt. Diese führten die Methode durch und bewerteten sie anschließend.

### **Methode:**

Zur Evaluation der Methode wurde ein auf HTML und Adobe Flash basierendes Bewertungstool entwickelt (vgl. Anhang J). Die in dieser Studie gewählte internationale Stichprobe bestand aus Ärzten, Pflegekräften, Experten der Ergonomie, der allgemeinen Mensch-Maschine-Interaktion und der medizinischen Mensch-Maschine-Interaktion (insb. aus dem für diese Normen zuständigen deutschen Spiegelgremium der DKE). Die Befragung wurde sowohl innerhalb eines persönlichen Interviews als auch online, verbunden mit einer Telefonkonferenz in deutscher und englischer Sprache, durchgeführt.

Im Rahmen der Befragungen sollten die Teilnehmer der Studie nach einer Einführung und Erklärung der Bewertungsmethode acht verschiedene Problemszenarien bewerten. Die Szenarien stellten Beschreibungen von Gebrauchstauglichkeitsproblemen medizinischer Geräte dar. Sie wurden sowohl aus der Literatur als auch aus den im Experimental-OP Tübingen aufgefallenen Gebrauchstauglichkeitsproblemen ausgewählt. Sie gliederten sich in eine Szenarienbeschreibung und einen Bewertungsbereich (Abb. 87). Innerhalb der Szenarienbeschreibung wurden dem Nutzer Informationen über das verwendete Gerät, die Probandengruppe des Gebrauchstauglichkeitstests, die zu lösende Arbeitsaufgabe, das Problemszenario, die Folgen des Problems während des Tests und die Auftretenshäufigkeit gegeben. Hierbei wurde darauf geachtet, dass die Beschreibungen neutral und wertfrei den Sachverhalt aus den Gebrauchstauglichkeitstests wiedergaben. Zusätzlich wurden in einigen Fällen die Szenarienbeschreibungen mit multimedialen Daten (Bilder, Video u. Audio) ergänzt.

Vor dem Test wurden die Teilnehmer über den Zweck der Evaluation aufgeklärt und für die Thematik sensibilisiert. Daraufhin bewerteten die Evaluatoren die einzelnen Elemente der Matrix (vgl. Anhang J). Als erstes wurde die Auftretens-

häufigkeit ermittelt. Hierbei musste der Anwender einschätzen, ob die Aufgabe eine Primär- oder Sekundäraufgabe ist, und ob diese „häufig“ oder „selten“ verwendet wird. Im Verhältnis zur Auftretenshäufigkeit im Gebrauchstauglichkeitstest wurde daraufhin die Auftretenshäufigkeit des Problems ermittelt. Im weiteren Verlauf sollte der Anwender das Erfüllen der Arbeitsaufgabe und somit das Problem sowie die Auswirkungen entsprechend der in Kapitel 6.1.1 beschriebenen Abstufungen bewerten. Zu jedem Bewertungsschritt konnten die Anwender eine Hilfe für weitere Informationen abrufen.

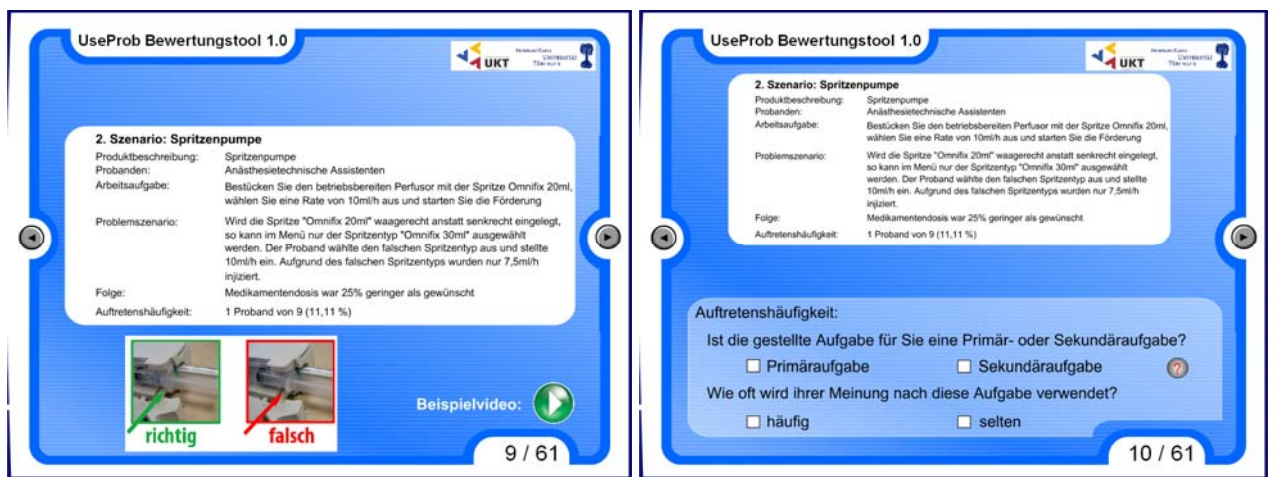


Abb. 87: Szenarienbeschreibung und Bewertungsbereich

Als erster Evaluationsschritt wurde nach jedem bewerteten Szenario dem Probanden das Ergebnis präsentiert und seine Zufriedenheit mit der Einstufung des Problems erfragt, oder, ob er die Einstufung zu positiv oder zu negativ empfand. Am Ende der Bewertung wurden verschiedene Fragen zur Zufriedenheit mit dem Konzept und dem Tool gestellt. Der für diese Studie entwickelte Fragebogen bestand aus drei Teilen.

Im ersten Teil wurden die demografischen Daten der Teilnehmer ermittelt. Im zweiten Teil wurde die Methode mit fünf geschlossenen Fragen anhand einer fünfstufigen Smiley-Skala und zwei Fragen mit verschiedenen vorgegebenen Antwortmöglichkeiten validiert (vgl. Anhang K). Die Fragen betrafen die Nutzermeinung zu der Methode insgesamt sowie zu einzelnen Elementen der Methode. Im dritten Teil konnten die Nutzer frei ihre positive und negative Meinung hinsichtlich der Methode äußern. Die Durchführbarkeit und Verständlichkeit der Methode wurde in mehreren Vorversuchen im Experimental-OP Tübingen geprüft. Der gesamte Fragebogen ist im Anhang K dargestellt.

Um zu ermitteln, welcher Personenkreis die Bewertung durchführen kann, wurden die potentiellen Evaluatoren analysiert und in Gruppen eingeteilt. Der in dieser Studie ermittelte Evaluatorenkreis stimmte weitgehend mit den von Kaye & Crowley identifizierten vier Personengruppen zur Bewertung medizinischer Risiken überein [212]. Neben klinischen Experten und Expertennutzern zählten auch Experten der Ergonomie, Gebrauchstauglichkeit und Mensch-Maschine-Kommunikation zu dem Evaluatorenkreis. Hierunter befanden sich auch professionelle Gutachter aus dem themenrelevanten Normungsgremium.

In dieser Studie wurde auf die Gruppe der Entwickler verzichtet, da davon auszugehen war, dass sie bzgl. ihres Produktes befangen sein würden und somit die Bewertung verfälschen könnten. Klinische Experten und Expertennutzer stellten in dieser Studie die Gruppe der Fachexperten dar. Innerhalb dieser Gruppe wurden Ärzte und Pflegepersonal unterschieden. Auch die Gruppe der Ergonomieexperten wurde in Untergruppen mit und ohne Erfahrungen im medizinischen Nutzungskontext unterteilt.

### Ergebnisse:

Insgesamt nahmen an der Studie 32 internationale Experten aus verschiedenen Fachrichtungen teil. Das Durchschnittsalter betrug 37,7 Jahre und die durchschnittliche Berufserfahrung in der Fachdisziplin 12 Jahre. Die Unterteilung der Experten in die einzelnen Fachrichtungen ist in Tab. 50 dargestellt.

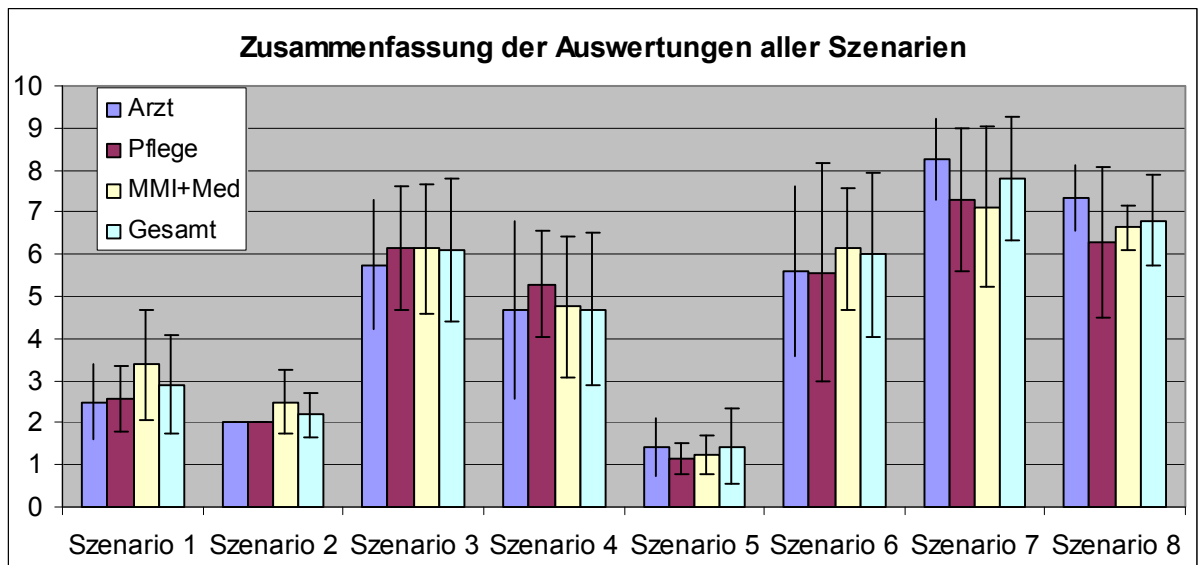
Fachgebiet	Teilnehmer	Ø-Alter	Ø-Berufserfahrung in Jahren
Insgesamt	32	37,7	12
Ärzte	10	12	38
Ärzte + Ergonomie	2		
Pflegekräfte	7	36,6	16,6
MMI-Experten	3	36	10,3
MMI- u. Medizin-Experten	9	42	13,9
Sonst. (Studenten)	2	24,5	0,25

Tab. 50: Teilnehmer der 2. Entwicklungsstudie

Bereits zu Beginn der Studie wurde in den durchgeführten Evaluationen und Gesprächen mit den Experten deutlich, dass Studenten und Experten, die lediglich Erfahrung in der Ergonomie und der Mensch-Maschine-Kommunikation besaßen, ohne über medizinisches Kontextwissen zu verfügen, Schwierigkeiten bei der Bewertung hatten. Daher wurden diese Stichprobengruppen nicht weiter einbezogen und die vorhandenen Ressourcen auf die drei weiteren Stichprobengruppen aufgeteilt, um diese zu verstärken. An der Studie nahmen ebenfalls auch

zwei Ärzte teil, welche Erfahrung in den Bereichen der Ergonomie und der Mensch-Maschine-Kommunikation hatten. Diese werden im Folgenden aufgrund ihres ausgeprägten medizinischen Fachwissens der Gruppe der Ärzte zugeordnet.

Abb. 88 veranschaulicht zusammenfassend die Ergebnisse der einzelnen Szenarien. Detaillierte Ergebnisse mit den jeweiligen Szenarienbeschreibungen werden im Anhang L aufgeführt.

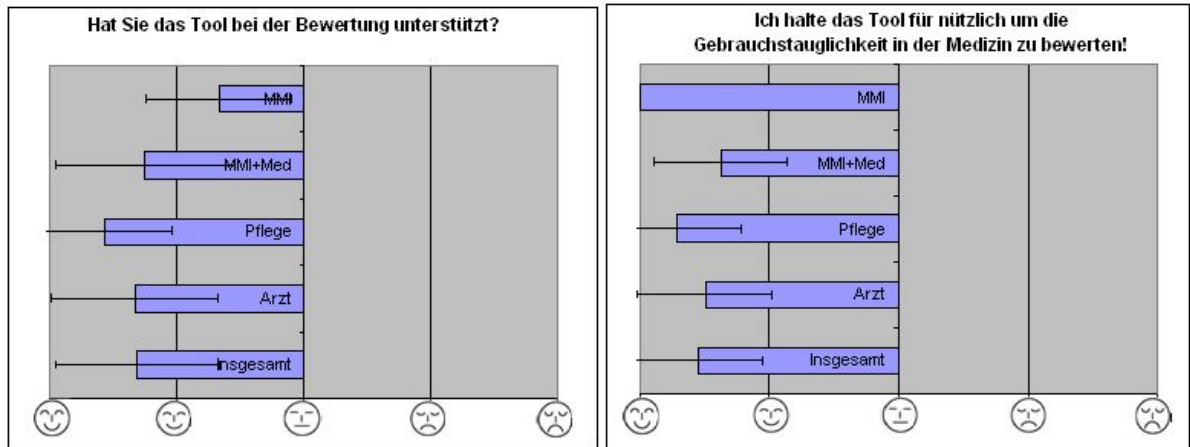


**Abb. 88: Zusammenfassung der Bewertung der evaluierten Szenarien mit Standardabweichungen**

Nach der Bewertung der einzelnen Szenarien war zu erkennen, dass die verschiedenen Evaluatorengruppen im Mittelwert der Bewertung eng zusammenlagen. Abweichungen zwischen den Bewertungen der Gruppen sind in der Regel kleiner als eine Bewertungsstufe (Abb. 88).

Weiterhin konnte sich die Bewertungsmethode in ihrer Verwendbarkeit und ihrem Nutzen als positiv herausstellen. Insgesamt war festzustellen, dass die Befragten das Bewertungstool durchweg positiv bewerteten und es als eine Unterstützung bei der Bewertung von Gebrauchstauglichkeitsproblemen in der Medizin ansahen (Abb. 89). Auch fanden sie die Anwendung des Tools einfach.





**Abb. 89: Expertenzufriedenheit mit der entwickelten Methode zur Bewertung der Gebrauchstauglichkeit medizinischer Produkte**

Alle Ergebnisse der subjektiven Zufriedenheitsmessung der einzelnen Fachexperten mit der Methode, welche im zweiten Evaluationsschritt ermittelt wurde, sind im Anhang M angegeben.

Insgesamt beschrieben die Evaluatoren aller Fachdisziplinen die Methode als einfach zu verwenden und geeignet für die Beurteilung der Gebrauchstauglichkeit medizinischer Geräte. Ein Auszug entsprechender Aussagen ist im Folgenden dargestellt:

- „Einfache Möglichkeit, um die Gebrauchstauglichkeit im medizinischen Kontext zu bewerten“ (MMI+Med)
- „Standardisierte Methode“ (MMI+Med, Arzt 2x)
- „Differenzierung orientiert sich am bekannten Ampelschema“ (MMI+Arzt)
- „Neuer Ansatz, der subjektive Eindrücke möglichst objektiv darstellen kann“ (Arzt)
- „Einfach, auch für einen Laien zu verwenden“ (Arzt)
- „Möglichkeit der Auswertung; Fehler und Probleme können bewertet werden. So etwas sollte mehr in der Praxis angewendet werden“ (Pflege)

Ferner gaben die Evaluatoren auch weiteres Verbesserungspotential an. So waren z.B. einige Begriffe aus dem Bereich der Gebrauchstauglichkeit für klinische Fachexperten zunächst unklar. Weiterhin wurde von einem Evaluator der Hinweis gegeben, dass die Einbettung der Szenarien in eine klinische Situation die Beurteilung der auftretenden Probleme erleichtern könnte. Für einige Teilnehmer war besonders am Anfang die Verwendung der Methoden ungewohnt. Da sie im späteren Verlauf diese als gebrauchstauglich beurteilten, schlugen sie vor,

besonders für neue Anwender der Methode am Anfang ein abstrahiertes Bewertungsbeispiel zu präsentieren. Weiterhin wurde bemängelt, dass es keine enthaltende Antwortmöglichkeit gab. Besonders Evaluatoren ohne klinischen Sachverstand hatten bei einigen Szenarien Schwierigkeiten, die Auswirkungen der Folgen zu bewerten. Eine Pflegekraft gab an, dass Probleme im medizinischen Arbeitsumfeld oft situationsabhängig sind, und dass gerade in Stresssituationen auch leichtere Probleme schwerwiegendere Auswirkungen haben können. Sie würde diesen Sachverhalt gerne bei der Bewertung berücksichtigen.

### **Diskussion:**

Die Bewertung von Gebrauchstauglichkeitsproblemen medizinischer Produkte wurde mit einer 3-dimensionalen Matrix durchgeführt. Hierzu wurden Bewertungseinstufungen des Risikomanagements mit denen der Gebrauchstauglichkeit verknüpft. Insgesamt zeigte sich die gewählte Methode als geeignet und einfach anzuwenden. Acht aus der Praxis stammende Gebrauchstauglichkeitsprobleme wurden mit der Methode bewertet. Der größte Teil der Evaluatoren, welche die Bewertung durchführten, war mit dem Bewertungsergebnis zufrieden. In der Regel lag der Anteil der zufriedenen Evaluatoren bei über 75%. Lediglich bei Szenario 6 lag er bei 72% und bei Szenario 7 bei 66%. In jedem Szenario waren mindestens 2/3 der Evaluatoren mit der Bewertung zufrieden. Auch die abschließende Evaluationsfrage nach der Zufriedenheit mit den Bewertungsergebnissen zeigt, dass 28 von insgesamt 32 Teilnehmern, d.h. 87,5% der Evaluatoren die Ergebnisse als genau richtig einstufen. Dieses Resultat zeigt, dass die Einstufungen für die Matrix und deren Aufbau richtig gewählt wurden. Eine wie in der Auswertung von Szenario 7 angesprochene Anpassung der Bewertungsstufen ist somit nicht sinnvoll. Für die Weiterentwicklung wird die Methode durch eine Anpassungsmöglichkeit ergänzt, mit der Evaluatoren nach der Ergebnispräsentation einer Bewertung diese entsprechend ihrer Meinung angleichen können, denn schließlich soll die Methode als Hilfsmittel einer Bewertung dienen und nicht als Ersatz für menschliche Logik.

Einige der medizinischen Fachexperten zeigten sich am Anfang der Bewertung von den Fachbegriffen der Methode irritiert, erlernten diese jedoch schnell. Um diesen Prozess zu verbessern, sollten vor der Bewertung die Begriffe erklärt und ein abstraktes, neutrales Übungsbeispiel durchgeführt werden.

Die multimediale Darstellung der Szenarien zeigte sich als sehr geeignet, um die Probleme zu präsentieren. Künftig sollte hierauf weiter großer Wert gelegt werden. Zusätzlich könnte es von Vorteil sein, wenn die Evaluatoren selbst die zu evaluierenden Geräte in Augenschein nehmen, um die Probleme nachzuvollziehen.

Vorschläge der Evaluatoren, dass die Einbettung der Szenarien in eine klinische Situation die Beurteilung der auftretenden Probleme erleichtern könnte, und Einwände, dass auch kleine Probleme in gewissen Situationen schwerwiegende Folgen haben können, werden bei der weiteren Methodenentwicklung nicht berücksichtigt, da die Szenarien in allgemeiner und neutraler Weise die Situation darstellen sollen. Eine Einbettung in Anwendungssituationen könnte die Ergebnisse durch Suggestionseffekte verfälschen.

Bei dieser Studie zeigte sich, dass Experten auf dem Gebiet der allgemeinen Mensch-Maschine-Interaktion Schwierigkeiten mit der Bewertung hatten. Hier war die größte Abweichung zu den restlichen Evaluatoren festzustellen. Daher wurden bereits am Anfang dieser Studie nur noch Experten der Mensch-Maschine-Interaktion mit Erfahrungen im medizinischen Bereich befragt. Bei allen anderen Evaluatorengruppen waren die Mittelwerte der Ergebnisse relativ nah beieinander. Dennoch waren in einigen Fällen gewisse Unterschiede feststellbar. Da es sich bei dieser Bewertung um eine subjektive Einschätzung der Fachexperten handelt, sind in der Regel unterschiedliche Meinungen vertreten. Relativ hohe Standardabweichungen, z.B. in Szenario 6, sind ein Beleg hierfür. Daher sollten unbedingt mehrere Evaluatoren eine Bewertung durchführen und der entsprechende Mittelwert gebildet werden [284]. Nielsen schlägt hierfür drei bis vier Experten vor [283, 284]. Diese Anzahl scheint auch für die Fragestellung dieser Arbeit sinnvoll zu sein und wird in der nächsten Evaluationsstufe überprüft. Auf jeden Fall sollten sowohl medizinische Fachexperten als auch Experten der medizinischen Mensch-Maschine-Interaktion (in dieser Arbeit auch „Doppelexperten“ genannt) die Bewertung durchführen, um den Erfahrungsschatz beider Fachbereiche miteinander zu verbinden.

Insgesamt erwies sich die gewählte Methode als sinnvoll und geeignet, um die Gebrauchstauglichkeit medizinischer Geräte zu bewerten. Sowohl Fachexperten der Medizin als auch Fachexperten der Mensch-Maschine-Interaktion bewerteten sie nicht nur positiv, sondern stufen sie als eine Bereicherung ein. Die in dieser

Studie erlangten Ergebnisse werden in die Methodenoptimierung einfließen und entsprechend dem iterativen Entwicklungsprozess weiter evaluiert.

Besonders interessant in dieser Studie war folgende Aussage eines Fachexperten aus dem Bereich der medizinischen Mensch-Maschine-Interaktion zur Einstufung der Gefährlichkeit eines Gebrauchstauglichkeitsproblems: „Das kann ich doch gar nicht beurteilen, ich bin doch kein Arzt“. Hierzu muss an dieser Stelle erwähnt werden, dass gerade Experten aus diesem weniger medizinisch orientierten Fachbereich derzeit die Gebrauchstauglichkeit medizinischer Geräte prüfen und entsprechende Verbesserungen herleiten. Aus dieser Diskrepanz wird in besonderem Maße klar, dass bei der Gebrauchstauglichkeitsevaluation unbedingt eine Bewertungsmethode eingesetzt werden sollte, welche den Erfahrungsschatz verschiedener Fachwelten miteinander verbindet.

### **6.2.3 Entwicklungsstudie 3 - Bewertung der HF-Geräte**

#### **Fragestellung:**

Ziel der dritten Entwicklungsstudie war die erste praktische Anwendung der Methode an einem medizinischen Gerät. Hierbei wurde überprüft, wie die Methode am besten einzusetzen sei. Fragestellungen hierbei waren die Festlegung der Evaluatorenanzahl und die Erörterung, ob die Bewertungsmethode in einer Einzelsitzung oder in einer Gruppensitzung (z.B. Fokusgruppe (vgl. Kap. 2.4.1.1)) angewandt werden sollte. Bei der Gruppensitzung wurde noch unterschieden, ob sich die Gruppe innerhalb einer Diskussion auf ein Ergebnis einigen sollte, oder ob die Gruppenmitglieder im Anschluss an die Diskussion die Geräte unabhängig voneinander zu bewerten hätten, um daraus schließlich einen Mittelwert bilden zu können.

#### **Material und Methode:**

Als Bewertungsgrundlage wurden die identifizierten Probleme der in Kapitel 4.2.1 beschriebenen Hochfrequenzchirurgiegeräte-Studie herangezogen. Die Probleme wurden den Teilnehmern vorgestellt, damit diese sie mit der Bewertungsmethode begutachten konnten. Hierbei wurden die durchgeführten Arbeitsaufgaben und die aufgefallenen Probleme erklärt und Highlightvideos aus den Nutzertests präsentiert. Die Bewertung erfolgte mit der in der Entwicklungsstudie 2 beschriebenen Methode.

Entsprechend der in Kapitel 4.1.3 ermittelten Erkenntnisse nahmen an dieser Studie vier Experten unterschiedlicher Fachrichtungen teil. Sie setzten sich aus zwei Ergonomieexperten mit Erfahrungen im medizinischen Bereich, einem Chirurgen und einer OP-Fachkrankenschwester zusammen.

Insgesamt wurde die Bewertung im Zeitraum von sechs Monaten dreimal durchgeführt. In der ersten Phase bewerteten die Experten in einer Einzelsitzung die Gebrauchstauglichkeitsprobleme der HF-Geräte. Durch Mittelwertbildung konnte eine erste Bewertung der HF-Geräte durchgeführt werden. In der zweiten Phase wurden in einer moderierten Gruppensitzung die Probleme diskutiert, wobei jedes für die Bewertung relevante Merkmal Berücksichtigung fand. Jeder der Experten sollte hierbei sein Fachwissen den anderen Experten vermitteln. Nach der Diskussion sollten sich die Experten auf ein Bewertungsergebnis einigen. In der dritten Gruppensitzung wurden in Anlehnung an die zweite Sitzung die Probleme diskutiert. Der Unterschied zur zweiten Sitzung war, dass die Teilnehmer unabhängig von den anderen Experten bewerten konnten. Hierzu führten sie nach der Problemdiskussion selbständig eine Bewertung durch. Aus den Ergebnissen wurde dann der Mittelwert berechnet.

### **Ergebnisse:**

Die Tab. 51 stellt die Ergebnisse der **Einzelbewertung** dar. Neben den Bewertungen der Experten, dem Mittelwert und der Standardabweichung werden die Abweichungen zwischen den Extrembewertungen aufgeführt. Von einer starken Abweichung wird gesprochen, wenn die Differenz zwischen zwei Extremwerten größer als drei ist. Dieser Bereich ist farblich hervorgehoben.

HF-Gerät 1							
Proble-m	Experte				Mittelwert	Sta. Abw.	Max Abw.
	1	2	3	4			
1	7	6	7	9	7,3	1,26	3
2	7	2	4	6	4,8	2,22	5
3	8	8	6	5	6,8	1,50	3
4	5	5	3	2	3,8	1,50	3
5	6	6	8	4	6,0	1,63	4
6	7	5	7	7	6,5	1,00	2
7	7	2	2	3	3,5	2,38	5
8	6	2	3	3	3,5	1,73	4
9	8	7	9	6	7,5	1,29	3
10	9	7	7	7	7,5	1,00	2
11	9	9	7	7	8,0	1,15	2
12	9	7	9	7	8,0	1,15	2
13	9	9	9	9	9,0	0,00	0
14	8	8	9	9	8,5	0,58	1
15	7	9	9	8	8,3	0,96	2
16	9	9	8	9	8,8	0,50	1
17	8	4	7	5	6,0	1,83	4
18	9	9	9	9	9,0	0,00	0
19	9	9	5	5	7,0	2,31	4
20	8	8	8	5	7,3	1,50	3
21	5	5	9	7	6,5	1,91	4
22	9	9	8	9	8,8	0,50	1
23	9	9	9	5	8,0	2,00	4
24	8	9	9	9	8,8	0,50	1
25	9	3	8	9	7,3	2,87	6
26	9	7	9	9	8,5	1,00	2
27	8	5	9	9	7,8	1,89	4

HF-Gerät 2							
Proble-m	Experte				Mittelwert	Sta. Abw.	Max Abw.
	1	2	3	4			
1	7	7	9	9	8,0	1,15	2
2	6	7	4	7	6,0	1,41	3
3	5	5	3	9	5,5	2,52	6
4	6	7	6	3	5,5	1,73	4
5	7	8	6	6	6,8	0,96	2
6	6	3	3	3	3,8	1,50	3
7	7	7	8	7	7,3	0,50	1
8	5	2	6	7	5,0	2,16	5
9	6	3	6	3	4,5	1,73	3
10	6	5	6	7	6,0	0,82	2
11	9	9	9	9	9,0	0,00	0
12	8	9	8	9	8,5	0,58	1
13	9	9	8	9	8,8	0,50	1
14	5	5	6	7	5,8	0,96	2
15	9	9	8	8	8,5	0,58	1
16	7	9	8	9	8,3	0,96	2
17	8	8	7	7	7,5	0,58	1
18	6	4	5	5	5,0	0,82	2
19	8	7	9	5	7,3	1,71	4
20	8	9	9	7	8,3	0,96	2
21	9	9	9	9	9,0	0,00	0
22	8	7	9	9	8,3	0,96	2
23	8	9	7	8	8,0	0,82	2
24	9	9	8	5	7,8	1,89	4
25	9	9	9	9	9,0	0,00	0

HF-Gerät 3							
Proble-m	Experte				Mittelwert	Sta. Abw.	Max Abw.
	1	2	3	4			
1	7	7	3	9	6,5	2,52	6
2	7	7	7	7	7,0	0,00	0
3	8	9	7	7	7,8	0,96	2
4	7	7	7	7	7,0	0,00	0
5	8	9	6	7	7,5	1,29	3
6	9	9	9	9	9,0	0,00	0
7	3	3	3	3	3,0	0,00	0
8	7	7	7	7	7,0	0,00	0
9	7	6	7	5	6,3	0,96	2
10	7	7	6	5	6,3	0,96	2
11	8	7	6	5	6,5	1,29	3
12	9	7	9	7	8,0	1,15	2
13	9	7	7	4	6,8	2,06	5
14	5	1	3	2	2,8	1,71	4
15	9	9	8	8	8,8	0,50	1
16	9	9	8	9	8,8	0,50	1
17	9	9	9	9	9,0	0,00	0
18	8	8	9	9	8,5	0,58	1
19	9	9	9	9	9,0	0,00	0
20	8	7	7	9	7,8	0,96	2
21	9	9	9	9	9,0	0,00	0
22	8	5	7	8	7,0	1,41	3
23	9	9	7	8	8,3	0,96	2
24	8	5	9	8	7,5	1,73	4
25	7	9	9	9	8,5	1,00	2
26	9	9	8	9	8,8	0,50	1

Tab. 51: Bewertungen der Gebrauchstauglichkeitsprobleme in der Einzelsitzung

Insgesamt lagen die Evaluatoren bei den meisten Problemen in der Bewertung nahe beieinander. In einigen Fällen wichen die Bewertungen jedoch stark voneinander ab. Bei der Bewertung des HF-Gerätes 1 war in 37% der Fälle die Abweichung zwischen der besten und der schlechtesten Bewertung höher als „3“. Bei der Bewertung des HF-Gerätes 2 traf dies in 20% und bei HF-Gerät 3 in 15% der Fälle zu.

In der **zweiten Bewertungsrunde** wurden die Probleme diskutiert und die Gruppe einigte sich auf eine Bewertung. Tab. 52 stellt die Ergebnisse der Gruppendiskussion dar. Als Vergleichswerte werden das Ergebnis der ersten Sitzung und der Betrag der Differenz zwischen den Sitzungen dargestellt. In dieser Sitzung wurden nur die Probleme diskutiert, welche in der vorherigen stark abwichen.

HF-Gerät 1				HF-Gerät 2				HF-Gerät 3			
Probl.	1. Sitzung	2. Sitzung	Differenz	Probl.	1. Sitzung	2. Sitzung	Differenz	Probl.	1. Sitzung	2. Sitzung	Differenz
1	7,3			1	8,0			1	6,5	7	0,5
2	4,8	3	1,8	2	6,0			2	7,0		
3	6,8			3	5,5	6	0,5	3	7,8		
4	3,8			4	5,5	6	0,5	4	7,0		
5	6,0	7	1	5	6,8			5	7,5		
6	6,5			6	3,8			6	9,0		
7	3,5	7	3,5	7	7,3			7	3,0		
8	3,5	6	2,5	8	5,0	6	1	8	7,0		
9	7,5			9	4,5			9	6,3		
10	7,5			10	6,0			10	6,3		
11	8,0			11	9,0			11	6,5		
12	8,0			12	8,5			12	8,0		
13	9,0			13	8,8			13	6,8	9	2,2
14	8,5			14	5,8			14	2,8	3	0,2
15	8,3			15	8,5			15	8,8		
16	8,8			16	8,3			16	8,8		
17	6,0	9	3	17	7,5			17	9,0		
18	9,0			18	5,0			18	8,5		
19	7,0	6	1	19	7,3	9	1,7	19	9,0		
20	7,3			20	8,3			20	7,8		
21	6,5	9	2,5	21	9,0			21	9,0		
22	8,8			22	8,3			22	7,0		
23	8,0	7	1	23	8,0			23	8,3		
24	8,8			24	7,8	9	1,2	24	7,5	3	4,5
25	7,3	8	0,7	25	9,0			25	8,5		
26	8,5							26	8,8		
27	7,8	9	1,2								

**Tab. 52: Bewertung der Gebrauchstauglichkeit durch Diskussion**

In der zweiten Sitzung wurde deutlich, dass die Bewertungen in den meisten Fällen überwiegend ähnlich ausfielen. Einige Bewertungsbeispiele zeigten jedoch, dass durch die zweite Bewertungsmethode auch erhebliche Abweichungen zu der ersten Bewertungsmethode zustande kamen. Dies ist bei der Bewertung des siebten Problems bei dem 1. HF-Gerät und des 24. Problems bei dem 3. HF-Gerät der Fall. Ein plausibler Grund hierfür könnte sein, dass durch die Diskussion der einzelnen Fachbereiche einigen Experten Erkenntnisse zugänglich gemacht werden konnten, über die sie vorher nicht verfügten. Mit diesem neu erworbenen Wissen konnte eine Bewertung durchgeführt werden. Es könnte aber auch sein, dass beispielsweise durch den Einfluss sehr dominanter Experten die Bewertung anderer Experten verfälscht wurde.

In der **dritten Bewertungsrunde** wurden die Gebrauchstauglichkeitsprobleme diskutiert. Danach beurteilten die Experten unabhängig voneinander das Problem. Die Bewertung erfolgte durch Mittelwertbildung. Die Tab. 53 zeigt die Ergebnisse.

HF-Gerät 1					
Pro-blem	Sitzung				
	1	2	3		
	MW	Wert	MW	Sta. Abw	Max Abw
1	7,3				
2	4,8	3	6,25	0,50	1
3	6,8				
4	3,8				
5	6,0	7	5,50	1,29	3
6	6,5				
7	3,5	7	6,75	0,50	1
8	3,5	6	6,00	0,82	2
9	7,5				
10	7,5				
11	8,0				
12	8,0				
13	9,0				
14	8,5				
15	8,3				
16	8,8				
17	6,0	9	9,00	0,00	0
18	9,0				
19	7,0	6	7,25	0,50	1
20	7,3				
21	6,5	9	8,00	0,00	0
22	8,8				
23	8,0	7	7,25	0,50	1
24	8,8				
25	7,3	8	8,00	0,82	2
26	8,5				
27	7,8	9	7,50	0,58	1

HF-Gerät 2					
Pro-blem	Sitzung				
	1	2	3		
	MW	Wert	MW	Sta. Abw	Max Abw
1	8,0				
2	6,0				
3	5,5	6	3,75	0,96	2
4	5,5	6	6,25	0,50	1
5	6,8				
6	3,8				
7	7,3				
8	5,0	6	6,00	0,00	0
9	4,5				
10	6,0				
11	9,0				
12	8,5				
13	8,8				
14	5,8				
15	8,5				
16	8,3				
17	7,5				
18	5,0				
19	7,3	9	8,50	1,00	2
20	8,3				
21	9,0				
22	8,3				
23	8,0				
24	7,8	9	8,25	0,96	2
25	9,0				

HF-Gerät 3					
Pro-blem	Sitzung				
	1	2	3		
	MW	Wert	MW	Sta. Abw	Max Abw
1	6,5	7	6,75	0,50	1
2	7,0				
3	7,8				
4	7,0				
5	7,5				
6	9,0				
7	3,0				
8	7,0				
9	6,3				
10	6,3				
11	6,5				
12	8,0				
13	6,8	9	8,75	0,50	1
14	2,8	3	3,25	0,50	1
15	8,8				
16	8,8				
17	9,0				
18	8,5				
19	9,0				
20	7,8				
21	9,0				
22	7,0				
23	8,3				
24	7,5	3	5,75	0,50	1
25	8,5				
26	8,8				

**Tab. 53: Bewertung der Gebrauchstauglichkeit durch Diskussion und anschließender unabhängiger Bewertung (MW=Mittelwert)**

In der dritten Sitzung war zu erkennen, dass die Bewertungen in der Regel nahe an den Ergebnissen der zweiten Sitzung lagen. Dennoch gab es Abweichungen, welche jedoch in der Regel nicht höher als zwei Bewertungsstufen waren. Innerhalb der Gruppe wurde Einheitlichkeit sichtbar. Im Gegensatz zur ersten Sitzung waren keine Abweichungen von mehr als drei Bewertungsstufen vorhanden, lediglich bei einem Problem kam eine Abweichung von drei Bewertungsstufen vor. Die Abb. 90 bis Abb. 92 veranschaulichen zusammenfassend die Ergebnisse (Anhang N) aller drei Sitzungen.



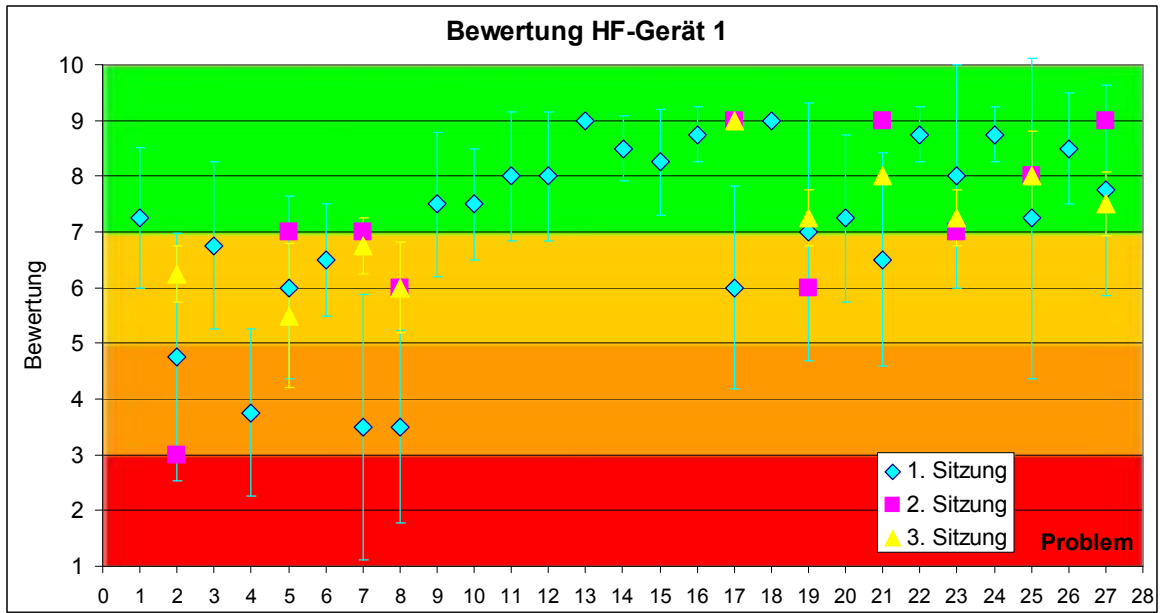


Abb. 90: Bewertung des 1. HF-Gerätes in drei Sitzungen. Bei der zweiten und dritten Sitzung wurden nur die Probleme mit hoher Abweichung bewertet.

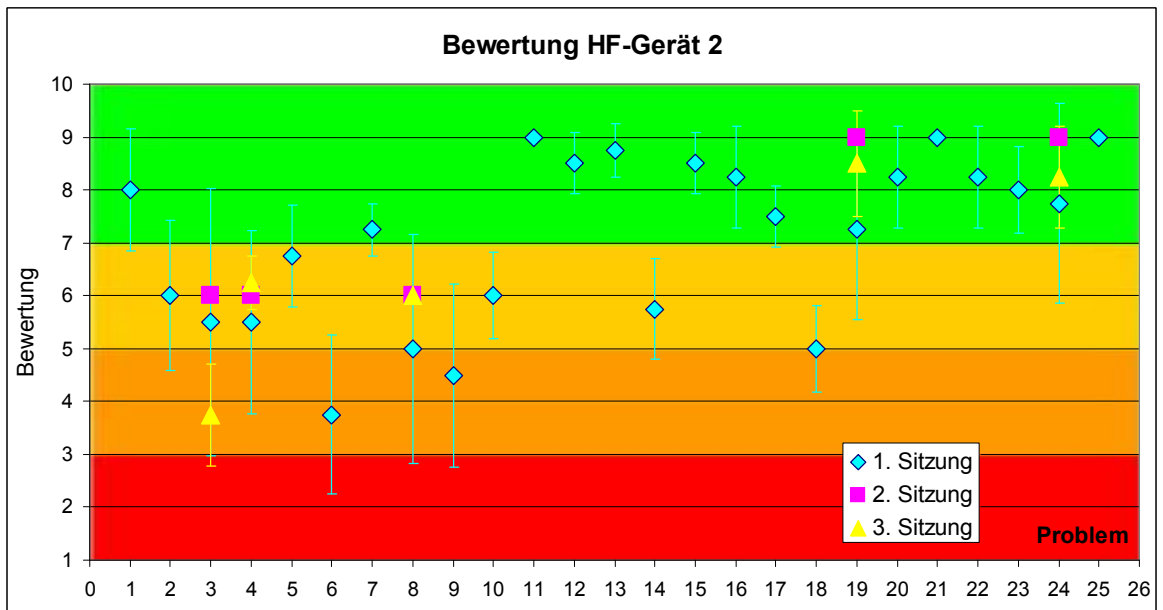
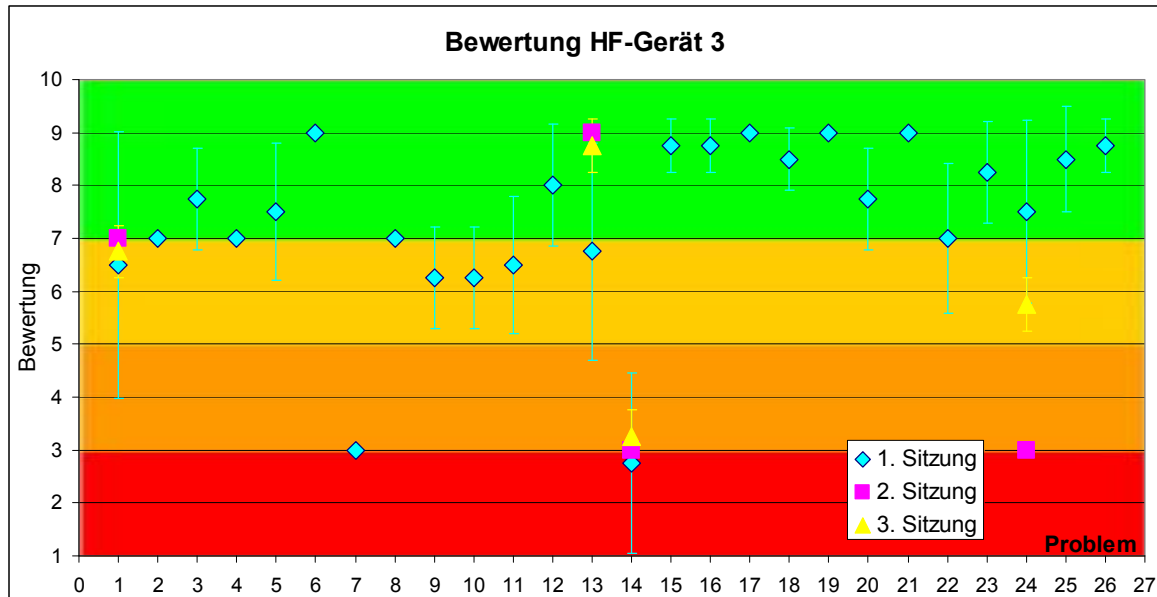


Abb. 91: Bewertung des 2. HF-Gerätes in drei Sitzungen. Bei der zweiten und dritten Sitzung wurden nur die Probleme mit hoher Abweichung bewertet.



**Abb. 92: Bewertung des 3. HF-Gerätes in drei Sitzungen. Bei der zweiten und dritten Sitzung wurden nur die Probleme mit hoher Abweichung bewertet.**

### Diskussion:

Die subjektive Bewertung richtet sich immer nach den subjektiven Einstellungen der Evaluatoren. Die Auffassungen von Menschen sind unterschiedlich, daher ist bei Bewertungen, wie sie in dieser Arbeit vorliegen, nicht nur die Meinung eines Experten zu berücksichtigen. Um die Reliabilität zu steigern, muss grundsätzlich die Anzahl der Experten erhöht werden. Die Ergebnisse der ersten Sitzung belegen diesen Sachverhalt. Zwar sind die Experten bei den meisten Problemen einer Meinung und beurteilen sie ähnlich, jedoch sind auch in einigen Fällen Abweichungen zu erkennen. Diese Abweichungen können jedoch nicht allein durch unterschiedliche Expertenmeinungen erklärt werden. Eine wichtige Rolle spielt auch, dass die Gebrauchstauglichkeit in der Medizin eine sehr interdisziplinäre Fragestellung ist, für deren Bearbeitung das Fachwissen verschiedener Fachdisziplinen benötigt wird. Experten, welchen das Fachwissen anderer Disziplinen fehlt, haben somit auch ein Wissensdefizit bei der Bewertung von Einzelfaktoren eines Problems. So fällt es z.B. einem Ergonomen meist schwer, die Auswirkungen eines Fehlers auf den Patienten zu beurteilen. Einem Arzt fällt es umgekehrt schwer, die Art eines Gebrauchstauglichkeitsproblems zu identifizieren und die Arbeitsbelastung des Anwenders zu bewerten. Daher müssen bei Gebrauchstauglichkeitsbeurteilungen die verschiedenen Kompetenzfelder zusammengebracht werden, damit ein Wissenstransfer stattfindet.

In der ersten Bewertungssitzung konnten die Experten nicht miteinander kommunizieren, daher ist diese Art der Bewertung ungeeignet. In der zweiten Bewertungssitzung diskutierten die Experten der einzelnen Fachdisziplinen miteinander und einigten sich auf eine Bewertung. Durch die stattfindende Diskussion erhielten die Evaluatoren Einblicke in das jeweilige Fachgebiet und konnten mit diesem Wissen die Probleme bewerten. Im Vergleich zur ersten Sitzung erkennt man, dass in einigen Fällen das Fachwissen bestimmter Teilnehmer unverzichtbar war, wodurch die Ergebnisse erheblich in eine andere Tendenz gerückt wurden. Problematisch war hierbei jedoch, dass sich die Evaluatoren am Ende der Diskussion auf ein Problem einigen sollten. Diese Vorgehensweise widerspricht der Tatsache, dass Individuen unterschiedliche Meinungen haben und diese auch vertreten. Zusätzlich kann es hierbei vorkommen, dass dominante Evaluatoren andere beeinflussen, und somit ihre Bewertung zu stark ins Gewicht fällt.

Besser geeignet ist die Methode der dritten Bewertungssitzung. Die Ergebnisse der mit ihr durchgeführten Evaluation zeigen nur noch geringe Abweichungen zwischen den Evaluatoren. Bei dieser Methode findet einerseits ein Meinungs- und Erfahrungsaustausch statt, andererseits werden verschiedene Meinungen zugelassen, sodass eine unabhängige Bewertung gewährleistet ist. Ein passendes Beispiel ist das Problem 24 bei der Bewertung des 3. HF-Gerätes: In der ersten Sitzung wurde dieses Problem als nicht sehr bedeutsam eingestuft (7,5), jedoch mit einer hohen Standardabweichung (1,73). In der zweiten Sitzung einigten sich die Experten auf eine Bewertung von 3. In der letzten Sitzung wurde eine Bewertung von 5,75 erzielt mit einer geringeren Standardabweichung von 0,5. Dies lässt vermuten, dass in der zweiten Sitzung ein zu dem Problem eher negativ eingestellter Experte seine Bewertung durchgesetzt hatte, obwohl andere Meinungen existierten.

Die Bewertung sollte daher durch eine interdisziplinäre Expertenrunde durchgeführt werden. Hierbei sollten mindestens zwei Vertreter der beteiligten Fachdisziplinen teilnehmen. Für den klinischen Fachbereich können die Experten nach dem Aufgabenspektrum unterschieden werden. Auch wäre es ratsam, innerhalb der Nutzergruppen sowohl einen Vertreter aus der Ärzteschaft als auch aus der Pflege als Experten einzusetzen, da diese meist unterschiedliche Aufgaben an den Geräten durchführen. Die gewählte Evaluatorengruppe von vier Experten zeigte sich als ausreichend für eine Bewertung. Es sollten jedoch nicht weniger zum Einsatz kommen. Grundsätzlich gilt: Je höher die Evaluatorenzahl, desto reliabler

und valider sind die Ergebnisse [36]. Sollten die Geräte von mehreren Anwendergruppen verwendet werden, so ist zusätzlich zu den Ergonomieexperten mindestens ein Vertreter der jeweiligen Anwendergruppe in die Bewertung mit einzubeziehen.

Treten bei der Bewertung weiterhin hohe Abweichungen auf (z.B. mehr als drei Bewertungsstufen), so ist eine weitere Bewertung und Konsultation zusätzlicher Experten zugunsten der Reliabilität ratsam. Optional könnte man zusätzlich einen Vertreter des Herstellers zu einer Produktbewertung einladen. Hierbei sollte man jedoch bedenken, dass dieser meist versuchen wird, das Gerät zu verteidigen. Andererseits kann er wichtige Informationen darüber liefern, warum bestimmte Produkteigenschaften gewählt wurden und welchen Nutzen diese haben.

In der Bewertung hat sich gezeigt, dass mit der Standardmethode und der Bewertungsmethode Gebrauchstauglichkeitsprobleme verschiedenster Schweregrade aufgedeckt und beurteilt werden können. Die Spannweite geht von kosmetischen bis hin zu potentiell sicherheitsgefährdenden Problemen. Im Folgenden wird die in dieser Studie als geeignet erschienene Bewertungsmethode an einem weiteren Praxisbeispiel evaluiert.

#### **6.2.4 Entwicklungsstudie 4 - Bewertung der Spritzenpumpe**

##### **Fragestellung:**

Ziel der vierten Entwicklungsstudie war es, die bisherigen Erkenntnisse zu evaluieren. Hierbei sollte die Bewertungsmethode verwendet werden, welche sich in der dritten Entwicklungsstudie als geeignet herausgestellt hatte. Zusätzlich sollte aus den einzelnen Problembewertungen ein Gesamtergebnis für das evaluierte Gerät berechnet werden.

##### **Material und Methode:**

Als Bewertungsgrundlage wurden die identifizierten Probleme der in Kapitel 4.2.3 beschriebenen Spritzenpumpen-Studie herangezogen. Die Probleme wurden den Teilnehmern vorgestellt, damit diese sie mit der oben genannten Methode bewerteten. Die Präsentation der Szenarien erfolgte wie in der dritten Entwicklungsstudie. Zusätzlich konnten die Experten die Geräte selbst begutachten, um die Arbeitsaufgaben durchzuführen und die Probleme zu reproduzieren. Bei der Bewertungsmethode wurde, entsprechend den Erkenntnissen der zweiten Entwicklungsstudie, eine Adaptionmöglichkeit eingeführt: Die Experten konnten, wenn

sie mit einem von dem Bewertungstool vorgegebenen Bewertungsvorschlag nicht einverstanden waren, diesen ihren Präferenzen gemäß angleichen.

Unter Berücksichtigung der bisherigen Erkenntnisse (vgl. Kap. 4.1.3) und im Einklang mit den Ergebnissen der Nutzungskontextanalyse (vgl. Kap. 4.2.3) nahmen an dieser Studie fünf Experten unterschiedlicher Fachrichtungen teil. Im Einzelnen waren dies zwei Ergonomieexperten mit Erfahrungen im medizinischen Bereich, ein Anästhesist sowie eine Anästhesie- und eine OP-Fachkrankenschwester.

Bisher wurden mit der Methode lediglich Probleme der Gebrauchstauglichkeit bewertet. Für die Bewertung eines medizinischen Gerätes reicht dies allein jedoch nicht aus. Hierfür müssen auch die positiven Aspekte betrachtet werden.

Grundsätzlich sollte bei der Feststellung eines sicherheitsrelevanten Gebrauchstauglichkeitsproblems die entsprechende Bewertung auf das ganze Gerät bezogen werden. Dies ist dadurch zu begründen, dass ein Gerät, unabhängig von seiner allgemeinen Handhabung, bei der Feststellung eines sicherheitsrelevanten Problems keine bessere Bewertung erreichen kann, als das sicherheitskritische Problem selbst. Weiterhin müssen für die Bewertung auf jeden Fall alle Primäraufgaben und ein Teil der Sekundäraufgaben in einem Nutzertest evaluiert und bewertet werden. Ist dies auf Grund der Komplexität des zu bewertenden Gerätes nicht möglich, so ist im Rahmen der Heuristischen Evaluation eine Auswahl der Aufgaben zu treffen.

Zur Bewertung der Arbeitsaufgabe sollte das schwerwiegendste Problem herangezogen werden [119, 281]. Zur Bewertung der Gebrauchstauglichkeit des gesamten Gerätes sollte der Mittelwert der bewerteten Aufgaben berechnet werden. Wenn jedoch ein sicherheitskritisches Problem vorhanden ist (Problemstufe 1-3), wird dieses für die Bewertung des ganzen Systems gelten, solange das Problem nicht beseitigt wurde. Da Nutzer hauptsächlich auf Primäraufgaben stoßen, sollten diese mit einem Faktor berücksichtigt werden. Dieser sollte individuell für jedes Gerät bestimmt werden. Ein möglicher Vorschlag ist der Faktor 4. Somit werden Bewertungsergebnisse von Primäraufgaben viermal stärker für die Bewertung des Gerätes berücksichtigt, als die von Sekundäraufgaben.

Die Formel für die Berechnung ist somit:

$$\text{Bewertungsergebnis} = \begin{cases} \text{MIN NP} & \text{wenn MIN NP} \leq \text{MIN NS} \\ & \text{und MIN NP} \leq 3 \\ \text{MIN NS} & \text{wenn MIN NS} \leq \text{MIN NP} \\ & \text{und MIN NS} \leq 3 \\ \frac{\sum_{i=1}^n NP_i * F + \sum_{j=1}^m NS_j}{n * F + m} & \text{in sonstigen Fällen} \end{cases}$$

MIN= Minimum   NP = Note Primäraufgabe   NS = Note Sekundäraufgabe  
F = Faktor   n = Anzahl der Primäraufgaben   m = Anzahl der Sekundäraufgaben

### Ergebnisse:

In dem Nutzertest wurden 14 Probleme gefunden, welche sich durch Abweichungen in Effektivität, Effizienz und Zufriedenheit bemerkbar machten. Mit Hilfe des durchgeführten Nutzertests und dessen Analyse konnten Problemschwerpunkte und -ursachen bei der getesteten Spritzenpumpe aufgedeckt werden (vgl. Kap. 4.2.3). Wie schwerwiegend die gefundenen Probleme sind, zeigte sich bei deren Bewertung. Den Experten wurden in einer moderierten Gruppensitzung die Probleme und Highlightvideos aus dem Nutzertest vorgeführt. Sie führten die Aufgaben selbstständig durch, um anschließend eine Bewertung abzugeben. Die entsprechenden Bewertungskriterien wurden in der Gruppe diskutiert. Auf der Basis ihrer Argumente nahmen die jeweiligen Experten die Bewertung vor. Die Bewertungsergebnisse wurden aus dem Mittelwert der fünf Bewertungen gebildet und sind in Abb. 93 dargestellt.

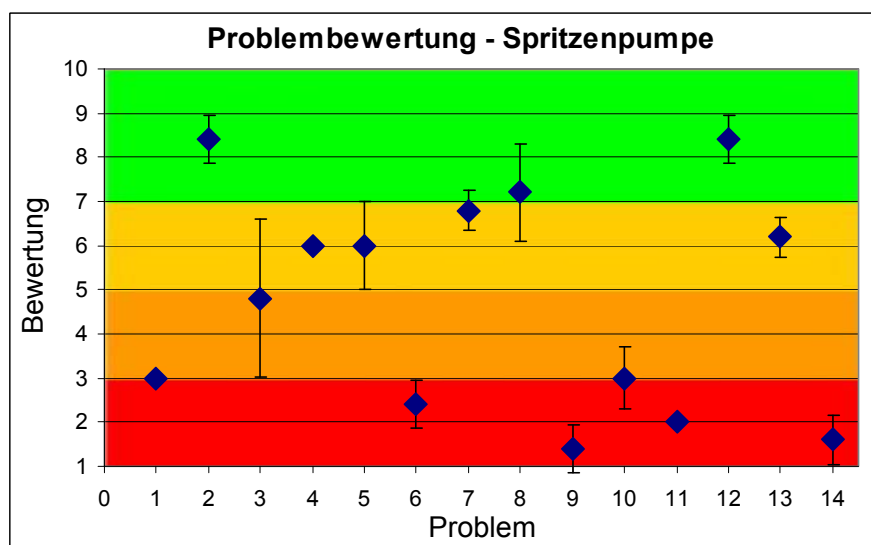


Abb. 93: Problembewertung durch fünf Experten mit Standardabweichung

Es zeigte sich, dass einige vermutete Probleme eher unbedeutend bzw. kosmetischer Art sind (grüner Bereich). Daneben gibt es verschiedene Probleme, welche die Gebrauchstauglichkeit beeinflussen und störend wirken können, jedoch keinen Einfluss auf die Sicherheit des Patienten haben (Abb. 93 gelb- und orange-farbiger Bereich). Aber es wurden auch vier Probleme entdeckt, welche einen Einfluss auf die Gesundheit des Patienten haben können und somit im nicht normkonformen Bereich mit der DIN EN IEC 60601-1-6 liegen (roter Bereich). Hierzu gehören (genaue Problembeschreibung siehe Anhang O):

- Problem 6: Die Spritze kann auch mit horizontal liegendem Spritzenstempel eingesetzt werden. Sie wird daraufhin von der Spritzenpumpe als eine Spritze mit anderem Volumen erkannt. Es kann so zu einer unbemerkten Fehldosierung kommen.
- Problem 9: Bei der Aufgabe, einen Bolus zu verabreichen, erfolgt dieser nicht, und die gesamte Förderung bricht ab.
- Problem 11: Ein Proband entlüftet durch „Bolusdrücken“ im Hauptmenü die Leitung und verabreicht dadurch unbemerkt einen erhöhten Bolus.
- Problem 14: Die Nutzer vermissen während der gesamten Bedienung ein Feedback über den aktuellen Status.

Die identifizierten und bewerteten Probleme ließen sich sechs Arbeitsaufgaben (vgl. Anhang P) zuordnen. Das Bewertungsergebnis für die Gebrauchstauglichkeit der Spritzenpumpe lag bei 1,4, da im Falle eines sicherheitskritischen Problems die Bewertung desselben für das ganze Gerät gilt.

### **Diskussion**

Bei der Bewertung erwies sich die Methode als geeignet. Alle Experten fanden sie hilfreich, um Gebrauchstauglichkeitsprobleme in der Medizin zu bewerten. Durch die multimediale Präsentation der Ergebnisse des Gebrauchstauglichkeitstests und die Selbsterfahrung durch eigene Interaktion mit den Geräten konnten die Experten die Probleme verstehen und deren Tragweite beurteilen. In der Diskussion der interdisziplinären Fachexperten konnten spezielle Fachkenntnisse ausgetauscht werden. Anschließend beurteilten die Experten unter Einbeziehung der neuen Gesichtspunkte die Ergebnisse unabhängig voneinander, woraus dann der Mittelwert gebildet wurde.

Bei den Bewertungsergebnissen traten lediglich geringe Abweichungen auf. Dies zeigt sich darin, dass in allen Fällen die Standardabweichung nicht größer als eine Bewertungsstufe ist. Lediglich bei dem dritten Problem war die Standardabweichung höher (1,79) (vgl. Anhang P). Dies erklärt, dass auch unterschiedliche Meinungen der Nutzer berücksichtigt werden. Fraglich ist hierbei, wie reliabel diese Ergebnisse sind. Dieser Fragestellung wird in der abschließenden Evaluation nachgegangen (► Kap. 6.4). Die Bewertung für das gesamte System betrug 1,4. Dieser Wert ist gerechtfertigt, da die Möglichkeit einer Patientengefährdung durch ein Gebrauchstauglichkeitsproblem gegeben ist.

### **6.3 Ergebnis - Bewertungsmethode „UseProb“**

In den durchgeführten Studien wurde eine theoretisch erarbeitete Bewertungsmethode für Gebrauchstauglichkeitsprobleme in der Medizin entwickelt und evaluiert. Diese Methode zur Bewertung medizinischer Gebrauchstauglichkeit wird von nun an „UseProb“ genannt. Im Folgenden werden die hieraus resultierenden Bewertungsschritte zusammengefasst.

#### **I. Problem- und Fehleranalyse**

Als vorbereitender Schritt für die Bewertung müssen Probleme und die hieraus entstehenden Fehler analysiert und kategorisiert (vgl. Kap. 5.2 u. 5.4) werden. Es muss geprüft werden, welcher Art der Fehler ist, und ob es sich dabei um einen durch ein Gebrauchstauglichkeitsproblem hervorgerufenen Fehler (vgl. Kap. 5.1.3), um einen rein menschlichen Fehler (vgl. Kap. 5.1.2), einen reinen Maschinenfehler oder um eine Mischform derselben handelt. Weiterhin müssen bei Gebrauchstauglichkeitsproblemen die problembeeinflussenden Faktoren für Effektivität, Effizienz und Zufriedenheit erfasst und dokumentiert werden. Hierzu zählen z.B. die Zeit, die benötigt wurde, um ein Problem zu beheben, der Grad der Beeinträchtigung der Zufriedenheit im Fehlerfall oder Auswirkungen eines Fehlers auf das Arbeitssystem.

#### **II. Zusammenstellung des Expertenteams**

Für die Bewertung muss ein interdisziplinäres Expertenteam zusammengestellt werden. Hierbei sollten sich die Erfahrungswerte der Einsatzbereiche des medizinischen Produktes und der Mensch-Maschine-Interaktion gegenseitig ergänzen.



Je mehr Evaluatoren eingesetzt werden, umso reliabler werden die Ergebnisse. Entsprechend den in dieser Arbeit gemachten Erkenntnissen sollten mindestens vier Experten für eine Bewertung herangezogen werden. Bei den medizinischen Experten sollte hierbei auf unterschiedliche Perspektiven geachtet werden. Aufgrund unterschiedlicher Arbeitsbereiche, Arbeitsaufgaben und Verantwortlichkeiten hat beispielsweise eine OP-Fachkrankenschwester durchaus eine andere Sichtweise als ein Chirurg.

### III. Bewertung

Die Bewertung sollte in einer moderierten Gruppensitzung stattfinden. Zuerst müssen die Gebrauchstauglichkeitsprobleme präsentiert und eventuell durch die Experten selbst nachgestellt werden. Weiterhin müssen den Experten die im Gebrauchstauglichkeitstest gemachten problemspezifischen Erkenntnisse veranschaulicht werden. Hierzu gehören beispielsweise die durchschnittliche Zeit zur Problembewältigung, Aussagen der Nutzer und Highlightvideos. Nachdem die Experten über die Probleme diskutiert haben, sollten sie diese mit Hilfe der in Kapitel 6.1.1 vorgestellten Methode bewerten. Aus den Ergebnissen wird der Mittelwert gebildet. Bei zu großen Abweichungen sollten gegebenenfalls weitere Experten konsultiert werden.

Zur Bewertung des medizinischen Gerätes ist folgende Formel zu verwenden:

$$\text{Bewertungsergebnis} = \begin{cases} \text{MIN NP} & \text{wenn MIN NP} \leq \text{MIN NS} \\ & \text{und MIN NP} \leq 3 \\ \text{MIN NS} & \text{wenn MIN NS} \leq \text{MIN NP} \\ & \text{und MIN NS} \leq 3 \\ \frac{\sum_{i=1}^n NP_i * F + \sum_{j=1}^m NS_j}{n * F + m} & \text{in sonstigen Fällen} \end{cases}$$

MIN= Minimum    NP = Note Primäraufgabe    NS = Note Sekundäraufgabe  
F = Faktor    n = Anzahl der Primäraufgaben    m = Anzahl der Sekundäraufgaben

### IV. Risikobetrachtung

Als letzte Maßnahme muss überprüft werden, ob das Risiko umgangen werden kann. Hierfür sollten mögliche Lösungswege für die Vermeidung des Gebrauchstauglichkeitsproblems gesucht werden. Sind Lösungsmöglichkeiten zur Vermeidung oder Verringerung des Risikos technisch oder ökonomisch nicht möglich, ist

zu prüfen, ob das in Kapitel 5.4.3 beschriebene ALARP-Prinzip greift. Ist ein durch ein Gebrauchstauglichkeitsproblem hervorgerufenen Risiko so niedrig wie vernünftiger Weise möglich, so darf es nicht in die Bewertung einfließen. Die Vertretbarkeit begründet sich durch die Bereitwilligkeit, mit dem Risiko aufgrund des gestifteten Nutzens zu leben.

## **6.4 Validierung - Finale Evaluation**

### **Fragestellung**

Zur Validierung der Standardprozedur wurden ihre beiden Entwicklungszweige zusammen evaluiert. Hierbei wurde überprüft, ob die entwickelten Methoden und die als geeignet befundenen Ansätze für eine Standardprozedur zur Überprüfung der Gebrauchstauglichkeit medizintechnischer Produkte zweckdienlich sind. Besonderes Augenmerk wurde hierbei auf die entwickelten Thesen und Anforderungen, sowie auf die Testgütekriterien Objektivität, Validität und Reliabilität (Kap. 4.1.8) gelegt. Die Studie soll zeigen, dass mit der in dieser Arbeit entwickelten Standardprozedur Gebrauchstauglichkeitsprobleme eines medizinischen Gerätes gefunden und bewertet werden können. Diese Ergebnisse müssen von anderen Prüfern, mit anderen Probanden und anderen Experten wiederholbar sein.

### **Material und Methode**

Als Untersuchungsgegenstände wurden eine, im Experimental-OP entwickelte Bedienschnittstelle für eine Herzkatheterpumpe der Firma CardioBridge gewählt sowie eines der Hochfrequenzchirurgiegeräte, deren Evaluation in Kapitel 4.2.1 beschrieben wurde.

Die Standardprozedur und die UseProb-Methode wurden zur Evaluation der Bedienschnittstelle angewandt. Damit die Testsituation realistischer wirkte und näher am Nutzungskontext war, wurden in dieser Studie die Aufgaben dem Testnutzer nicht in Form eines Aufgabenblattes vorgegeben, sondern in verbalisierter Form eines Rollenspiels. Der Testleiter nahm hierbei im Rahmen des Nutzungskontextes die Rolle eines Operators ein. Hierdurch sollte eine natürlichere Nutzungssituation erreicht werden. Weiterhin wurde der ASQ nicht wie bisher nach jeder Aufgabe gestellt, sondern nach Aufgabenblöcken. Hierdurch wurde vermieden, dass der Testteilnehmer nach jeder Aufgabe aus der Nutzungssituation geholt werden musste, um den Fragebogen zu beantworten.

Zur Messung der Testgütekriterien (vgl. Kapitel 4.1.8) wurden unabhängig voneinander zwei Testreihen mit derselben Benutzerschnittstelle durchgeführt. Die zweite Testreihe wurde von einer Psychologiestudentin und einem Doktoranden der medizinischen Fakultät der Universität Tübingen absolviert.

Die Ergebnisse beider Studien geben Hinweise über die Testgüte der Standardprozedur und über den in Kapitel 4.1.3 beschriebenen Evaluatoreffekt. Zusätzlich wurde ein zweiter Expertenworkshop zur Bewertung der Probleme bei den Hochfrequenzchirurgiegeräten und zur Ermittlung der Retestreliaibilität durchgeführt.

#### Messung der Objektivität

Objektivität drückt die Unabhängigkeit eines Verfahrens aus. Im Fall der Standardprozedur ist dies die Unabhängigkeit des Evaluationsleiters. In der Entwicklungsstudie 4 (vgl. Kap. 6.2.4) wurde bereits gezeigt, dass die Standardprozedur von unterschiedlichen Testleitern durchgeführt werden kann. In dieser Studie soll gezeigt werden, dass unterschiedliche Testleiter zu ähnlichen Ergebnissen gelangen. Hierbei wird die Durchführungsobjektivität geprüft. Somit müssen zwei unabhängige Testleiter ähnliche Probleme einer Schnittstelle aufdecken. Zur Ermittlung der Auswertungsobjektivität müssen die Probleme ähnlich bewertet werden. Zur Messung der Objektivität werden die gefundenen Fehler und die Bewertungen miteinander verglichen. Die Korrelation der Ergebnisse der verschiedenen Testleiter sollte eine Reliaibilität ( $r$ ) höher als 0,8 erreichen. Aufgrund der in Kapitel 4.1.8 beschriebenen Problematik der Reliaibilitätsmessung wird zusätzlich die Bland-Altman-Methode [29] angewendet.

#### Messung der Reliaibilität

Die Prüfung der Reliaibilität ist im Rahmen dieser Studie für die Problemfindung nicht möglich, da hierfür der gleiche Test mit derselben Stichprobe durchgeführt werden müsste (vgl. Kap. 4.1.8). Hierbei müssten dieselben Ergebnisse erzielt werden. Wird jedoch für einen Nutzertest zweimal dieselbe Stichprobe verwendet, so wird diese beim zweiten Test die Bedienschnittstelle schon kennen, und bestimmte Probleme treten nicht mehr in Erscheinung. Anders ist dies bei der Problembewertung. Diese ist unabhängig von der Geräteerfahrung.

In dieser Studie wird die Retestreliaibilität gemessen. Hierzu führt eine Expertengruppe zweimal mit einem zeitlichen Abstand die Bewertung durch. Zur Überprüfung der Reliaibilität muss nach wiederholter Anwendung bei derselben Stichprobe die Methode zu den gleichen Ergebnissen führen. Als Maß gilt die Korrelation der

beiden Messwertreihen (vgl. Kap. 4.1.8). Um von einem reliablen Test zu sprechen, sollte eine Korrelation von über  $r=0,75$  erreicht werden. Zusätzlich wird die Bland-Altman-Methode angewendet.

Für die Reliabilitätsmessung wurde ein Hochfrequenz-Chirurgiegerät aus der in Kapitel 4.2.1 beschriebenen Studie ausgewählt. Der zeitliche Abstand zwischen den beiden Bewertungssitzungen betrug sieben Monate.

#### Messung der Validität

Die Validität wurde durch eine Konsensbildung überprüft. Hierbei wurde die Standardprozedur durch verschiedene Experten dahingegen geprüft, ob die in Kapitel 4.1.8 beschriebenen Validitätskriterien eingehalten wurden.

#### Verifikation der Thesen

Als letzter Schritt der Validierung wurden die in Kapitel 4.1 aufgestellten Thesen auf ihren Wahrheitsgehalt überprüft. In diesem Zusammenhang dienten die Ergebnisse der einzelnen Studien dazu, diese zu be- oder zu widerlegen.

### **Ergebnisse**

In der Aufgabenanalyse wurden bei der zu bewertenden Bedienschnittstelle 17 Arbeitsaufgaben identifiziert. Diese unterteilten sich in sieben Primäraufgaben und zehn Sekundäraufgaben. Für die erste Studie wurden zwölf und für die zweite Studie 13 Testaufgaben gewählt (vgl. Anhang Q). Zehn der 13 insgesamt ausgewählten Aufgaben wurden in beiden Studien durchgeführt. Die Testaufgaben sind mit Bezug auf die in der Aufgabenanalyse identifizierten Aufgaben in Tab. 54 dargestellt. Man erkennt, dass hierbei die Primäraufgaben zu 100% übereinstimmen. Die Aufgabe „Geschwindigkeit verändern“ wurde in der zweiten Studie nicht einzeln durchgeführt. Die hier vorkommenden Aktionen sind jedoch auch in der Aufgabe „RCP starten“ enthalten. Die Aufgabe „Warnmeldung“ wurde in der zweiten Studie zweimal mit jeweils einer anderen Warnmeldung durchgeführt. Von den Sekundäraufgaben wurden nur vier in beiden Studien absolviert. Insgesamt wurden in beiden Studien 72% der identifizierten Aufgaben durchgeführt, da die erste Studie zwei Aufgaben in einer kombinierte, welche die zweite Studie getrennt prüfte.

Identifizierte Aufgaben in der Aufgabenanalyse	Teilaufgabe	Testaufgabe in Studie 1	Testaufgabe in Studie 2
<b>Setup</b>	Anschluss der Schläuche	1	1
	Drucknormierung	1	1
	Anschluss des Motors	1	1
	Funktionstest	1	1
	Einführung des Katheters	1	1
<b>RCP starten</b>	Öffnen	2	2
	Starten	2 + 8	2 + 7
	Geschwindigkeit einstellen	2 + 8	2 + 7
Spüllösung verändern		10	5
<b>Warnmeldung</b>	Temperatur	5	4
	Geschwindigkeit		9
<b>Geschwindigkeit verändern</b>		4	
Technischer Status			8
Verlauf			
Alarm-Verlauf		6	10
Helligkeit		3	3
Kontrast			
Lautstärke			
Service			
Hilfe		11	11
Tastensperre		9	
<b>Weaning</b>		12	12
<b>Notaus</b>		7	6
<b>Ausschalten</b>		12	13

Tab. 54: Gewählte Testaufgaben der identifizierten Aufgaben; Die Zahlen geben die Reihenfolge im Test an; Primäraufgaben sind fett dargestellt.

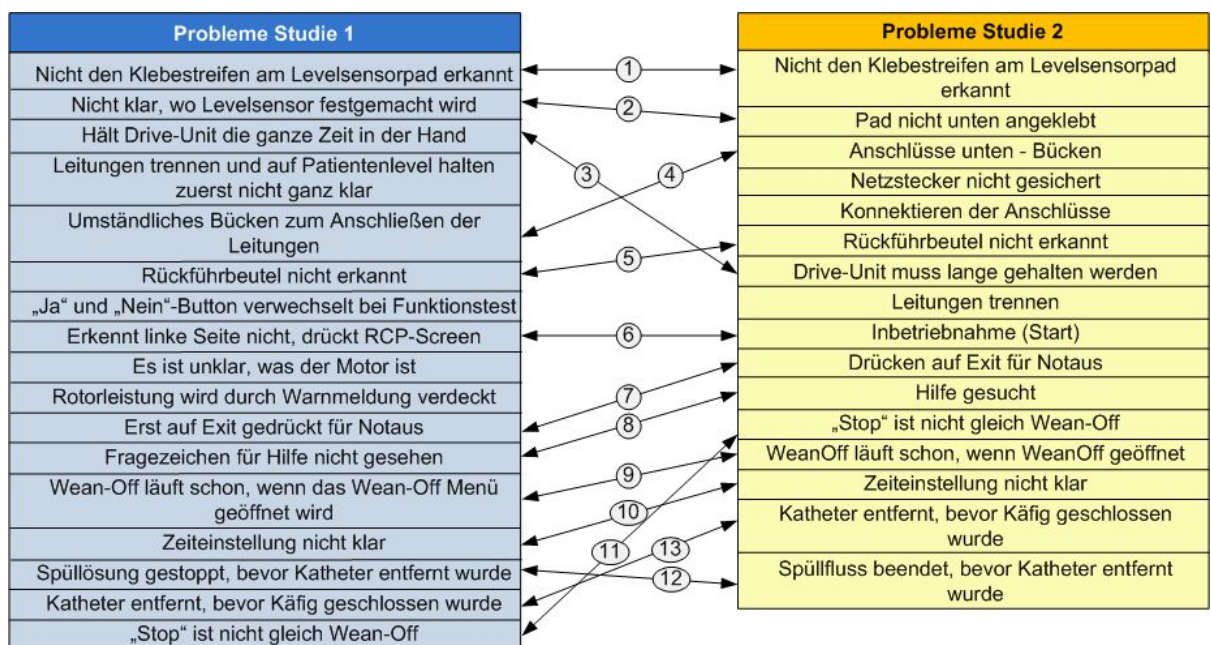
### Ergebnisse der Objektivitätsmessung

Die Stichprobengröße wurde in der ersten Studie mit neun und in der zweiten mit acht Teilnehmern angesetzt. Die **Durchführungsobjektivität** wurde durch einen Vergleich der in beiden Tests gefundenen Probleme gemessen. Nach Reduktion von Doppelnennungen und Nutzerwünschen sowie einer Zusammenfassung von Problemen konnten in beiden Nutzertests 42 Gebrauchstauglichkeitsprobleme identifiziert werden, welche sich durch Abweichungen in Effektivität, Effizienz und Zufriedenheit bemerkbar machten. 22 Probleme wurden hiervon in beiden Studien und 20 in nur einer Studie identifiziert (vgl. Anhang R). Dies entspricht einer prozentualen Übereinstimmung der Probleme von 52%. Reduziert man die Problemzahl auf die Probleme, welche in den Aufgaben auftraten, die in beiden Tests gestellt wurden, so konnten 22 von 39 Problemen in beiden Studien identifiziert werden (56%). Nachdem die Testverantwortlichen die 20, resp. 15 für sie wichtigsten Probleme heraussuchen sollten, entsprach für die 20 wichtigsten

Gebrauchstauglichkeitsprobleme die Übereinstimmung zu 68% und für die 15 wichtigsten zu 73%.

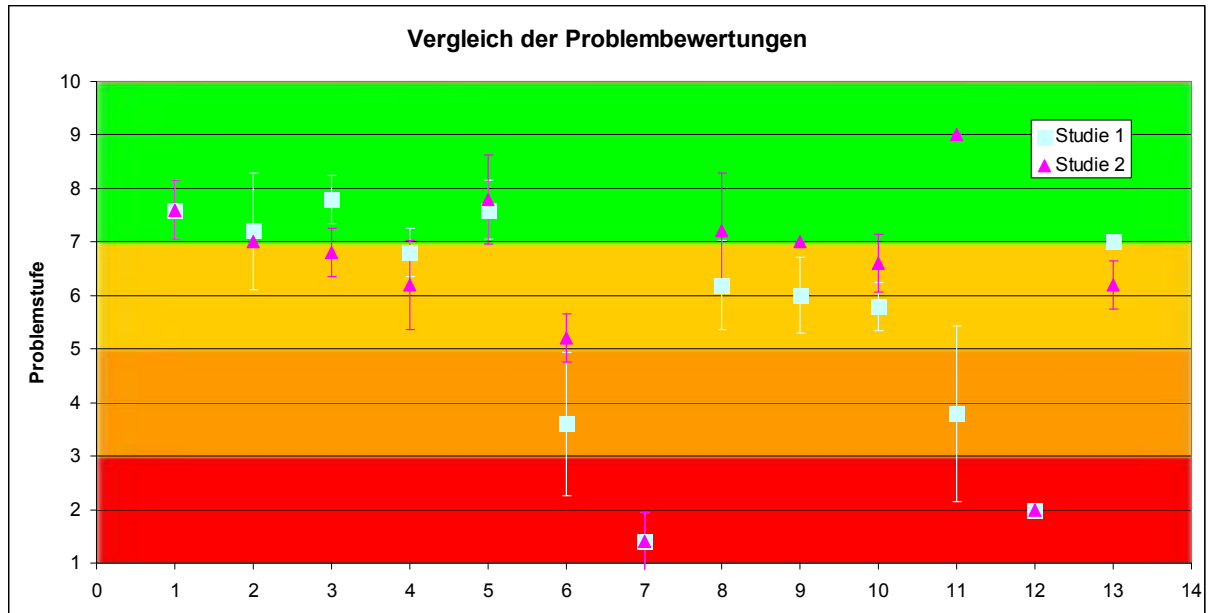
Die **Auswertungsobjektivität** wurde durch einen Vergleich der Problembewertungen ermittelt. Der Bewertungsworkshop wurde in beiden Studien mit fünf Teilnehmern durchgeführt. An dem ersten Bewertungsworkshop nahmen ein Chirurg mit Erfahrungen in der Ergonomie, ein Kardiologe, eine Fachkrankenschwester für Intensivmedizin, ein Experte für Mensch-Maschine-Kommunikation und ein Entwickler des Herstellers teil. Die Experten für den zweiten Workshop waren eine Psychologiestudentin mit Erfahrungen in der Mensch-Maschine-Kommunikation, ein Experte für Mensch-Maschine-Kommunikation, ein Kardio-techniker, eine OP-Fachkrankenschwester und ein zweiter Entwickler des Herstellers.

In der, der Bewertung vorausgehenden, Problemanalyse stellten sich einige Probleme als technische, andere als menschliche Probleme heraus (vgl. Kap. 5.2). Bei der Problembewertung wurden nur die Gebrauchstauglichkeitsprobleme ausgewählt. Auch sollte die Bewertung nicht durch zu viele Probleme in die Länge gezogen werden, daher wurde die Bewertung auf eine angemessene Anzahl beschränkt. Abb. 94 zeigt die bewerteten Probleme der beiden Studien und die Übereinstimmungen.



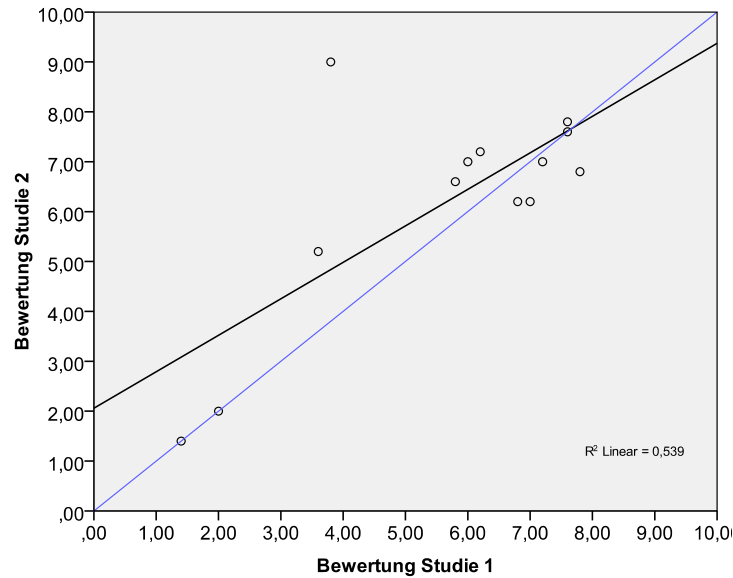
**Abb. 94: Ermittelte Gebrauchstauglichkeitsprobleme aus zwei getrennten Studien am gleichen Produkt (mit Nummerierung der ausgewählten Probleme und Übereinstimmungshinweisen).**

In der ersten Studie wurden 17 und in der zweiten Studie 16 Probleme bewertet. 13 dieser Probleme wurden in beiden Studien bewertet. Abb. 95 zeigt die Bewertungsergebnisse, welche in beiden Bewertungen generiert wurden. Die Nummerierung der Probleme entspricht den Problemnummern aus Abb. 94. Alle Bewertungsergebnisse sind dem Anhang P zu entnehmen.



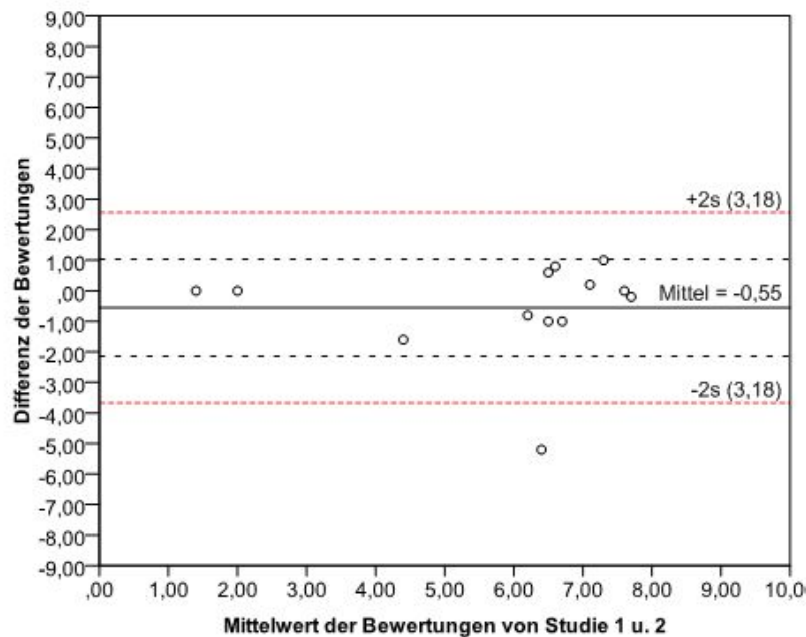
**Abb. 95: Problembewertung zweier Studien im Vergleich (vgl. Abb. 94) mit Standardabweichung**

Die Bewertungsergebnisse lagen bei den Bewertungen sehr nah beieinander. Die durchschnittliche Abweichung zwischen den Bewertungen betrug 0,9 Bewertungsstufen. Die meisten Bewertungen wichen nicht mehr als eine Bewertungsstufe voneinander ab. Lediglich die Bewertung von Problem 6 wich um 1,6 und die von Problem 11 um 5,2 Bewertungsstufen ab. Die Korrelation der Bewertungsergebnisse nach Pearson betrug 0,734 mit einer Signifikanz kleiner 0,01 (Anhang T).



**Abb. 96: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug auf die Bewertungsergebnisse der ersten Studie (x-Achse), mit der Regressionsgeraden ( $r=0,734$ ) und der Geraden der Gleichheit (Perfekte Korrelation,  $r=1$ , blau)**

Abb. 96 veranschaulicht die Ergebnisse der beiden Methoden und die Gerade der perfekten Korrelation. Abb. 97 veranschaulicht die Ergebnisse der Bland-Altman-Methode. Der Systematische Fehler ist der Mittelwert der Messwert-differenzen ( $\bar{d}$ ) und beträgt  $-0,55$ . Bei einer Normalverteilung liegen 95% der Werte in einem Bereich von  $\pm 3,18$  ( $\bar{d} \pm 1,96s$ ) um die Messwertdifferenz.

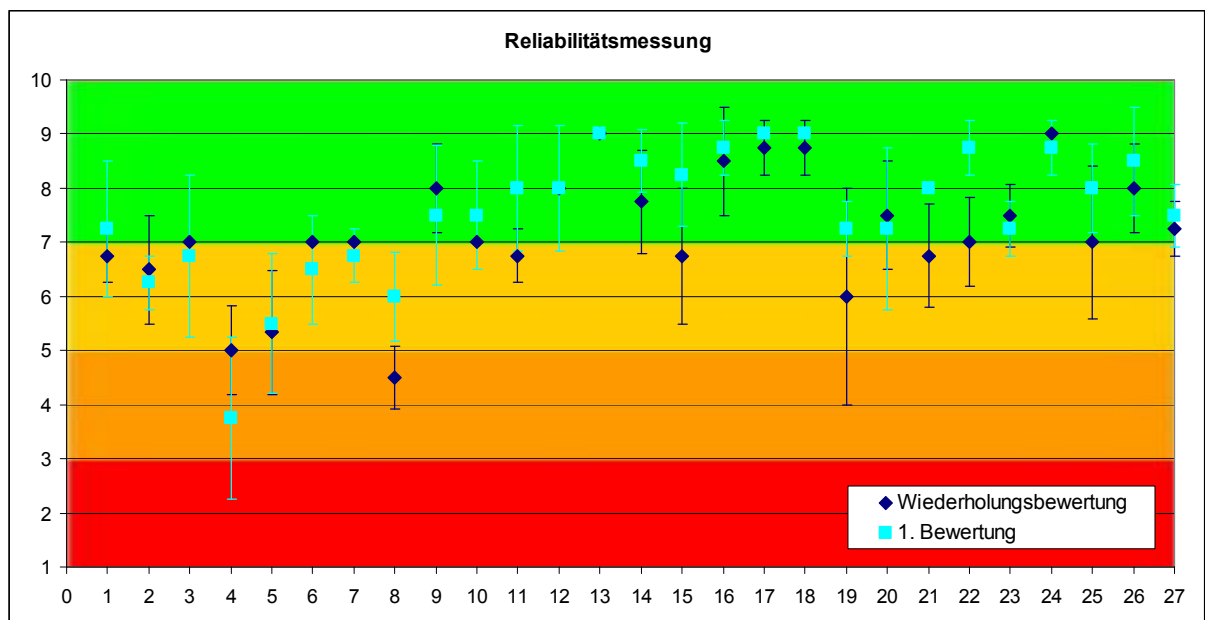


**Abb. 97: Bland-Altman-Methode zur Visualisierung des Zusammenhangs zweier wiederholender Studien**



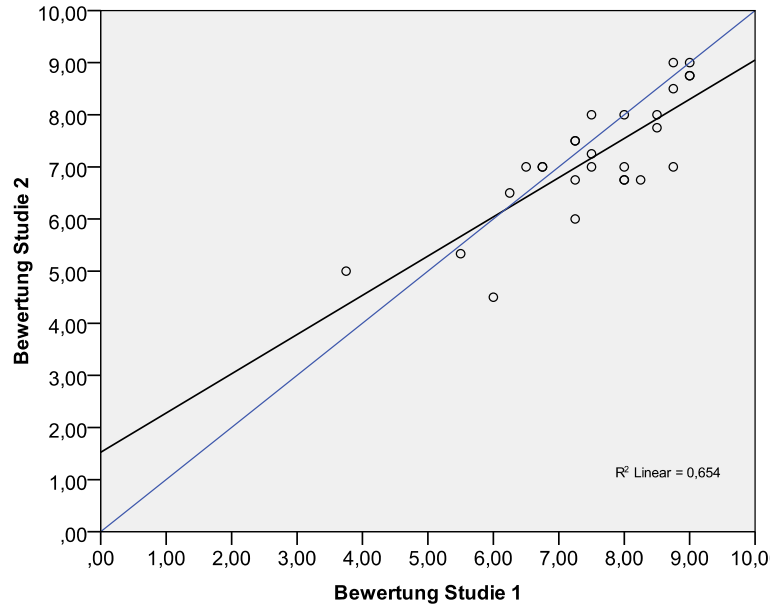
### Ergebnisse der Reliabilitätsmessung

Zur Messung der Reliabilität wurde die Bewertung der Probleme des ersten HF-Gerätes (vgl. Kap. 4.2.1) ein zweites Mal mit den gleichen Evaluatoren durchgeführt. Die Ergebnisse der Bewertung aus dem ersten Expertenworkshop und der zweiten Bewertung sind in Abb. 98 dargestellt. Hierbei war die maximale Abweichung der Probleme 1,75 (Problem 21). In 70% der Fälle war die Bewertung geringer als eine Bewertungsstufe und in 67 % der Fälle kleiner oder gleich einer halben Bewertungsstufe. Die durchschnittliche Abweichung betrug 0,6 Bewertungsstufen. Die Korrelation der Bewertungsergebnisse nach Pearson betrug  $r=0,81$  mit einer Signifikanz kleiner 0,001. Demnach kann von Reliabilität gesprochen werden. Die einzelnen Bewertungsergebnisse sind im Anhang U dargestellt.

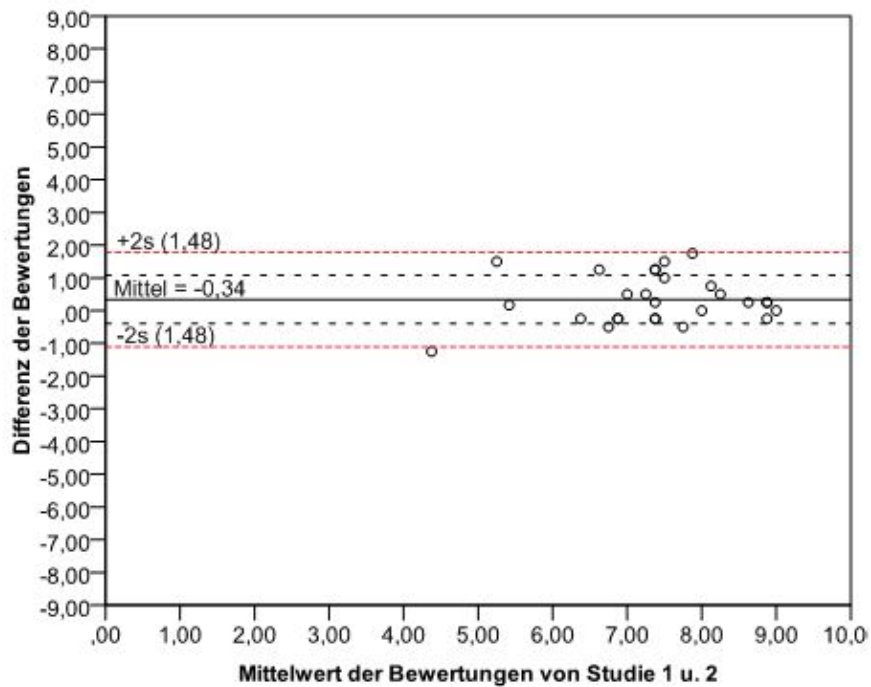


**Abb. 98: Ergebnisse der Problembewertung zweier wiederholender Studien mit Standardabweichung**

Abb. 99 veranschaulicht die Ergebnisse der beiden Methoden und die Gerade der perfekten Korrelation. Abb. 100 veranschaulicht die Ergebnisse der Bland-Altman-Methode. Der Systematische Fehler ist der Mittelwert der Messwert-Differenzen ( $\bar{d}$ ) und beträgt 0,339. Bei einer Normalverteilung liegen 95% der Werte in einem Bereich von  $\pm 1,476$  ( $\bar{d} \pm 1,96s$ ) um die Messwertdifferenz.



**Abb. 99: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug auf die Bewertungsergebnisse der ersten Studie (x-Achse), mit der Regressionsgeraden ( $r=0,81$ ) und der Geraden der Gleichheit (Perfekte Korrelation,  $r=1$ , blau)**



**Abb. 100: Bland-Altman-Methode zur Visualisierung des Zusammenhangs zweier wiederholender Studien**

**Gesamtbewertung:**

Die Gesamtbewertung der Katheterpumpe wurde mit der in Kapitel 6.3 beschriebenen Bewertungsformel berechnet. In den beiden Studien beträgt die Bewertung der Gebrauchstauglichkeit des Gerätes 1,4. Diese Einstufung entsteht dadurch, dass ein sicherheitskritisches Problem die Bewertung bestimmt. Würden die sicherheitskritischen Probleme nicht existieren, so würde die erste Studie ein Bewertungsergebnis von 7,8 und die zweite Studie ein Ergebnis von 8,1 ermitteln. Die Abweichung zwischen den beiden Bewertungsergebnissen beträgt somit 0,3 Bewertungseinheiten. Gewichtet man die Primäraufgaben mit dem Faktor 4, so sind die Bewertungen in der ersten Studie 7,2 und die in der zweiten 7,7. Auch hier beträgt die Abweichung lediglich eine halbe Bewertungsstufe. Die genaue Berechnung der Bewertungen ist im Anhang V dargestellt.

**Ergebnisse der Thesenverifikation**

Zur Überprüfung der im Rahmen dieser Arbeit aufgestellten und theoretisch hergeleiteten Gedankengänge wurden die in Kapitel 4.1 aufgestellten Thesen verifiziert. Hierzu wurden die relevanten Ergebnisse der Studie den entsprechenden Thesen zugeordnet. Die Verifikation erfolgte im Rahmen einer Diskussion. Auf Grundlage der Thesenverifikation wurde als letzter Schritt der finalen Evaluation der Erfüllungsgrad der in Kapitel 3.3 und Kapitel 6.2 spezifizierten Anforderungen an die Standardprozedur verifiziert. Hierzu wurden die Ergebnisse aller Studien, theoretische Hintergründe, die Erfahrungen aus der Methodenentwicklung und die Ergebnisse der Thesenverifikation herangezogen. Aus dieser abschließenden Verifikation sollte hervorgehen, ob die Entwicklung der Standardprozedur erfolgreich war. Die Ergebnisse werden im Folgenden dargestellt:

T.1	Der medizinische Nutzungskontext ist durch Benutzer, Arbeitsaufgabe, Arbeitsmittel und Arbeitsumgebung bestimmt.
	Bei allen in dieser Arbeit evaluierten medizinischen Geräten wird die Gebrauchstauglichkeit durch den Benutzer, die Arbeitsaufgabe, die Arbeitsmittel und die Arbeitsumgebung bestimmt. Zusätzlich könnte in gewissen Fällen die Gebrauchstauglichkeit noch speziell durch den Patienten oder die Systemintegration der medizinischen Geräte beeinflusst werden. Diese Parameter können aber weitestgehend dem Bereich „Arbeitsumgebung“ zugeordnet werden.

T.2	Für die Überprüfung der Gebrauchstauglichkeit medizinischer Geräte ist die Erfassung der beeinflussenden Faktoren des Nutzungskontextes unerlässlich.
<p>In den im Rahmen dieser Arbeit durchgeführten vier Entwicklungsstudien erweist sich die Erhebung des Nutzungskontextes in der Vorbereitungsphase als wichtig, um Informationen zu erhalten, die benötigt werden, um Gebrauchstauglichkeitsziele aufzustellen, Testaufgaben zu entwickeln, ein Profil für die Testnutzer zu erarbeiten und um die Testumgebung mit den Testszenarien zu gestalten.</p>	
T.3	Methoden zur Entwicklung des Nutzerprofils sind Literaturrecherche, Beobachtung, Nutzerinterview, Fragebogenbefragung, Hersteller- und Vertreiberinformationen oder Experteninterviews.
<p>Zur Erhebung des Nutzungskontextes können Literaturrecherche, Nutzerinterviews, Fragebogenbefragungen, Hersteller- und Vertreiberinformationen oder Experteninterviews verwendet werden, jedoch ist keine Methode so ergiebig wie eine Beobachtung des Nutzers in seiner natürlichen Arbeitssituation. Auch sollten mehrere Methoden angewandt werden. In Kapitel 4.2.1 wird gezeigt, dass beispielsweise die Angaben von Herstellern und Anwendern gravierend voneinander abweichen.</p>	
T.4	Eine Ermittlung und Unterteilung der Nutzergruppen ist notwendig.
<p>Eine Unterteilung der Nutzergruppen ist auf jeden Fall sinnvoll, da unterschiedliche Nutzergruppen oft unterschiedliche Aufgaben an Geräten durchführen. Hierbei ist sowohl nach Nutzern für spezielle Arbeitsaufgaben als auch nach unterschiedlichen Erfahrungen zu unterscheiden. Auch unerfahrene Nutzer sollen in einem Gebrauchstauglichkeitstest berücksichtigt werden.</p>	
T.5	Unterschiedliche Nutzer medizinischer Geräte benutzen diese anders und unterschiedliche Ergebnisse werden evaluiert.
<p>In allen Studien ist zu erkennen, dass bei unterschiedlichen Nutzern unterschiedliche Probleme auftreten. Es ist jedoch nicht davon auszugehen, dass unterschiedliche Nutzertypen zwingend unterschiedliche Probleme haben. So sind beispielsweise bei der Evaluation der minimalinvasiven Handgriffe (vgl. Kap. 4.2.2) die Ergebnisse zwischen Experten und unerfahrenen Nutzern signifikant unterschiedlich, wobei dieser Effekt bei der Evaluation der Spritzenpumpen (vgl. Kap. 4.2.3) oder der Herzkatheterpumpe (vgl. Kap. 6.4) nicht zu beobachten ist.</p>	

T.6	Als Methoden für ein Aufgabenprofil kommen Befragungen (Interviews, Fragebögen) von Nutzern und Herstellern, Beobachtungen des Anwenders, Brainstorming-Techniken, Fokusgruppen und/oder Kontextbegehungen in Betracht.
Befragungen der Hersteller liefern Erkenntnisse über die Aufgaben, welche mit dem Gerät durchgeführt werden können. Befragungen der Anwender und Kontextbegehungen liefern Einsicht in die Aufgaben, die tatsächlich von den Nutzern durchgeführt werden.	
T.7	Mit der Aufgabenanalyse können die mit dem System durchgeführten Ziele, Aufgaben und Aktionen der Nutzer ermittelt und kategorisiert werden. Sie liefert somit elementare Informationen für die Gebrauchstauglichkeitsprüfung.
Die Aufgabenanalyse stellt sich bei allen Evaluationen als zwingend notwendig heraus, um das zu untersuchende System zu erfassen und für den Nutzertest relevante Aufgaben auszuwählen. Die HTA (vgl. Kap. 4.1.1.2) erweist sich hierbei als übersichtlich und geeignet.	
T.8	Aufgaben sollten für die Evaluation in Primär- und Sekundäraufgaben unterteilt werden.
Für eine Standardisierung ist eine derartige Differenzierung wichtig. Die Standardprozedur verwendet diese Unterteilung, um mehr Konsistenz in der Aufgabenwahl für einen Gebrauchstauglichkeitstest zu erreichen. Die Bewertungsmethode verwendet diese Unterteilung für die Definition der Auftretenshäufigkeit eines Problems.	
T.9	Die Analyse der Arbeitsumgebung und der Arbeitsmittel liefert elementare Informationen für den Aufbau der Gebrauchstauglichkeitsprüfung.
In gut untersuchten Bereichen können eine Literaturrecherche und eine Kontextbegehung ausreichen.	
T.10	Die Umgebungsanalyse liefert Informationen darüber, welche Faktoren als leistungsbeeinflussend bei der Gebrauchstauglichkeitsevaluation berücksichtigt werden müssen.
Allein eine Analyse der Arbeitsumgebung lässt erkennen, welche Faktoren im Nutzungskontext einen Einfluss auf die Gebrauchstauglichkeit haben können. Hierzu zählen z.B. Stress, Beleuchtungssituationen oder Teamwork.	

T.11	Aus einer Nutzeranalyse, den Anforderungen an das Gerät und einer Expertenevaluation können Testkriterien abgeleitet werden.
<p>Erst nachdem der Nutzungskontext erfasst worden ist, erkennt der Testleiter, welche Gebrauchstauglichkeitsziele gesetzt und welche Anforderungen an das Produkt gestellt werden. Eine Expertenevaluation liefert Informationen über mögliche Schwachstellen eines Gerätes. Hieraus können messbare Kriterien für die Gebrauchstauglichkeitsbestimmung entwickelt werden.</p>	
T.12	Nutzertest und Heuristische Evaluation decken die meisten Probleme auf.
<p>In allen Gebrauchstauglichkeitsevaluationen dieser Arbeit hat sich eine Kombination der Gebrauchstauglichkeitsmethoden Heuristische Evaluation und Nutzertest bewährt. In der Spritzenpumpen-Studie wird keine Heuristische Evaluation angewandt. In der hier verwendeten unstrukturierten Expertenevaluation werden scheinbar weniger Probleme gefunden (vgl. Kap 4.2.3).</p> <p>In dieser Arbeit wird jedoch davon ausgegangen, dass die in einer Expertenevaluation vermuteten Probleme nicht immer wirkliche Probleme sein müssen. Daher wird diese Evaluationsmethode als Vorbereitung für den Nutzertest verwendet. Die hier vermuteten Probleme sollten im Nutzertest erhärtet werden.</p>	
T.13	Die Heuristische Evaluation sollte für medizinische Geräte angepasst werden.
<p>In den durchgeführten Studien zeigt sich, dass die Leitsätze der Heuristischen Evaluation nach Nielsen für medizinische Experten nicht eindeutig verständlich waren. Daher ist eine Anpassung auf das Vokabular dieses Expertenkreises sinnvoll. Zusätzlich sollte in der Heuristischen Evaluation der in der DIN EN 60601-1-6 beschriebene versehentliche Missbrauch betrachtet werden.</p>	
T.14	Die Heuristische Evaluation erfasst potentielle Probleme, welche im Gebrauchstauglichkeitstest überprüft werden können. Hierzu sollten entsprechende Aufgaben gewählt werden.
<p>Diese Vorgehensweise stellt sich in allen Studien als geeignet heraus. Somit können die in der Expertenevaluation vermuteten Probleme verifiziert und auf ihre Relevanz für den Nutzer überprüft werden.</p>	

T.15	Zur Messung der Gebrauchstauglichkeit in der Medizin werden Leistungsmaße und subjektive Maße benötigt.
Zusätzlich sollte jedoch auch die für die Medizin elementare Sicherheit beachtet werden. Bei mechanischen Tätigkeiten und potentiellen körperlichen Belastungen sollten auch Beanspruchungsmessungen durchgeführt werden (vgl. Kap. 4.2.2).	
T.16	Leistungsmaße können in Zusammenarbeit mit Expertennutzern aufgestellt werden.
Dies ist jedoch nur möglich, wenn Expertennutzer existieren. Bei Geräten, die noch nicht auf dem Markt sind, können Daten der Mitbewerber verwendet werden. Sollten diese auch nicht vorhanden sein, müssten diejenigen Personen Daten generieren, welche sich am längsten mit dem Gerät beschäftigt haben (Entwickler, Testleiter etc.) (vgl. Kap. 6.4).	
T.17	Für die Erhebung subjektiver Zufriedenheitsdaten können Fragebögen eingesetzt werden. Mit einem aufgabenspezifischen Fragebogen (z.B. ASQ), speziellen Fragen zu den einzelnen Hauptaufgaben an dem Gerät und der Erfassung des Gesamteindrucks ist eine Bewertung der Gebrauchstauglichkeit möglich.
Diese Vorgehensweise bestätigt sich in den Gebrauchstauglichkeitsevaluationen. Durch diese können sowohl Informationen über die Nutzerzufriedenheit zu dem gesamten System als auch zu einzelnen Systemelementen und Interaktionsfolgen gewonnen werden.	
T.18	Um die Gebrauchstauglichkeit in der Medizin zu bewerten, reichen die Maße Effektivität, Effizienz und Zufriedenheit nicht aus, es müssen zusätzlich die Aspekte Sicherheit bzw. Risiko betrachtet werden.
Bereits durch die Betrachtung des Nutzungskontextes zeigt sich, dass Sicherheit bei der Gebrauchstauglichkeit medizinischer Geräte ein wesentliches Attribut ist. Dennoch ist Sicherheit kein Maß für Gebrauchstauglichkeit, denn diese wird in erster Linie durch die Maße Effektivität, Effizienz und Zufriedenheit erreicht. Für die Betrachtung der die Gebrauchstauglichkeit betreffende Sicherheit wird daher in dieser Arbeit eine spezielle Bewertungsmethode entwickelt.	

T.19	Die Standardprozedur hilft, den Evaluatoreffekt zu vermeiden, da sie strukturiert den Nutzungskontext und die Ziele und Aufgaben analysiert, verschiedene Evaluatoren mit einbezieht und einen standardisierten Evaluationsleitfaden vorgibt.
Durch die Standardprozedur wird ein Grad der Standardisierung erreicht, welcher den Evaluatoreffekt einschränkt. In der finalen Evaluationsphase wird dies anhand der Testgütekriterien bewiesen (vgl. Kap. 6.4).	
T.20	Die Heuristische Evaluation sollte mit Doppelexperten und Fachexperten durchgeführt werden.
Doppelexperten haben sowohl Erfahrung in der Gebrauchstauglichkeitsbetrachtung als auch im medizinischen Kontext. Da jedoch das medizinische Arbeitsfeld weit aufgestellt ist, können Doppelexperten nicht für alle Disziplinen vorhanden sein, bzw. sind schwer zu akquirieren. Daher sollten auch Fachexperten eingesetzt werden.	
T.21	Die Bewertung sollte mit mehreren Doppel- und Fachexperten durchgeführt werden.
Durch den Einsatz von mehreren Doppelexperten und interdisziplinären Fachexperten werden verschiedene Sichtweisen in die Bewertung einbezogen. Die Ergebnisse werden so valider.	
T.22	Die Stichprobe sollte ein ausgewogenes Verhältnis von Laien zu Fachnutzern aufweisen.
In der Spritzenpumpen- und der Herzkatheterstudie (vgl. Kap. 4.2.3 u. 6.4) können keine gravierenden Unterschiede zwischen verschiedenen Nutzergruppen ermittelt werden. In der Gebrauchstauglichkeitsstudie bzgl. der minimalinvasiven Handgriffe werden dagegen unterschiedliche Ergebnisse zwischen Novizen und Laien festgestellt (vgl. Kap. 4.2.2). Dies zeigt, dass unbedingt auf ein ausgewogenes Verhältnis zwischen Fachnutzern und Laien geachtet werden sollte, da es durchaus sein kann, dass diese mit unterschiedlichen Problemen konfrontiert werden.	
T.23	Für die Stichprobe sind Probanden zu nehmen, welche dem Nutzerprofil entsprechen.
Nur eine ausgewogene Stichprobe führt zu einer kompletten Betrachtung des medizinischen Gerätes. Auf jeden Fall sollten auch Novizen mit entsprechendem Vorwissen einbezogen werden.	



T.24	Die Anzahl der Nutzer muss bei der Gebrauchstauglichkeitsprüfung in der Medizin höher als 6 sein. Die Problemfindungsrate sollte bei mindestens 90% liegen.
In den Studien zeigt sich, dass die Stichprobe höher als sechs Teilnehmer anzusetzen ist, um eine Problemfindungsrate von über 90% zu erreichen. In der finalen Studie wird eine Problemfindungsrate von 90% erst mit 13 Nutzern erreicht (vgl. Kap. 6.4).	
T.25	Mit einer Stichprobengröße von 9 Nutzern werden mehr als 80% der Probleme gefunden. Hiermit ist eine Bewertung möglich.
In den ersten Studien wird schon mit wesentlich weniger Testnutzern eine Problemfindungsrate von über 80% erreicht. In der finalen Studie (Kap.6.4) liegt dieser Wert jedoch nur knapp bei 80,5%. Daher sollten eher 10 Testnutzer für eine Bewertung eingesetzt werden.	
T.26	Anwendungsszenarien helfen bei der Definition des Nutzungskontextes und der Erstellung der Testaufgaben.
Anwendungsszenarien können bei der Definition des Nutzungskontextes helfen, vor allem, wenn noch kein realer Nutzungskontext existiert.	
T.27	Aufgabenszenarien vermitteln Nutzern in natürlicher Weise das Aufgabenziel und Informationen für die Aufgabe.
Dies wird erfolgreich in den ersten Studien angewandt (vgl. Kap. 4.2.1 - 4.2.3). In der Studie zur Gebrauchstauglichkeitsevaluation der Katheterpumpe zeigt sich jedoch, dass verbal formulierte Aufgabenszenarien in natürlichen Handlungsabläufen mehr Vorteile haben, da dies einer natürlichen Nutzungssituation entspricht. Einem Bias (Messfehler) durch unterschiedliche Aufgabenverbalisierungen sollte hierbei entgegengewirkt werden. Hierzu kann ein Drehbuch helfen.	
T.28	Aus den Anwendungsszenarien, den Kontextszenarien und den Informationen der Anforderungsanalyse können die für die Evaluation notwendigen repräsentativen Testaufgaben abgeleitet werden.
Diese Vorgehensweise bestätigt sich in den Gebrauchstauglichkeitsevaluationen. In allen Studien werden daher die Testaufgaben in diesem Sinne ausgewählt.	

T.29	Testaufgaben sollen hauptsächlich Primäraufgaben enthalten. Es sollen aber auch Sekundäraufgaben und Aufgaben in Notsituationen gestellt werden sowie Aufgaben, welche in der Anforderungsanalyse oder in einem Expertenreview als potentiell problematisch oder risikoträchtig identifiziert wurden. Zusätzlich soll auch, entsprechend den Forderungen der DIN EN 60601-1-6, ein vernünftigerweise vorhersehbarer Missbrauch betrachtet werden.
Mit den nach der Standardprozedur ermittelten Testaufgaben kann eine Bewertung der Gebrauchstauglichkeit medizinischer Geräte durchgeführt werden. Dies wird durch die Sicherstellung der Validität in Kapitel 6.4 gezeigt	
T.30	Gebrauchstauglichkeitsuntersuchungen in der Medizin müssen in einer Laborumgebung durchgeführt werden, in welcher der Nutzungskontext simuliert werden kann.
Tests sollten aus ethischen Gründen nicht im realen Nutzungskontext durchgeführt werden. In einer Laborumgebung ist die notwendige Aufnahme- und Beobachtungstechnik vorhanden, es sollte hier jedoch auf eine möglichst reale Abbildung des Nutzungskontextes geachtet werden.	
T31	Der für Softwareprodukte standardisierte Gebrauchstauglichkeitsbericht kann in einer angepassten Form für Medizinprodukte verwendet werden.
Die angepasste Form des Gebrauchstauglichkeitsberichts nach der ISO 25026 hat sich als Dokumentationsstandard in dieser Studie bewährt.	
T.32	Die Standardprozedur erfüllt die klassischen Testgütekriterien.
Die Standardprozedur erfüllt die Testgütekriterien Objektivität, Reliabilität und Validität. Sollten bei der Bewertung eines Problems zu starke Abweichungen auffallen, dann sollten zusätzliche Fachexperten hinzugezogen werden.	
T.33	Objektivität wird durch die Standardprozedur erreicht. Die Korrelation zwischen zwei Testanwendern sollte bei einem Wert größer 0,8 liegen.
Die Standardprozedur ist von Einstellungen der Bewerter und einem Informationsaustausch innerhalb der Bewertungsgruppe abhängig. Durch die Methode wird in dieser Studie eine relativ hohe Korrelation von $r=0,73$ erreicht. Dies entspricht zwar nicht dem erwarteten Wert, dennoch wird von einer Objektivität ausgegangen. Die entsprechende Diskussion hierzu ist Kapitel 6.4 zu entnehmen.	

T.34	Eine hohe Reliabilität kann durch die Standardprozedur erreicht werden. Ein Test nach wiederholter Anwendung führt zu den annähernd gleichen Ergebnissen ( $r=0,75$ ).
Reliabilität wird in dieser Arbeit durch die Standardprozedur erreicht. Ein Test nach wiederholter Anwendung führt zu den annähernd gleichen Ergebnissen. Die Korrelation beträgt $r=0,81$ .	
T.35	Zumindest bei den schwerwiegendsten Problemen sollte in einem wiederholten Test die gleiche Benotung herauskommen.
Die Probleme, welche als sicherheitskritisch bewertet worden sind, erhalten in wiederholten Tests eine ähnliche Bewertung. Abweichungen des Mittelwerts der Problembewertung waren nicht feststellbar (vgl. Problem 7 und 12 Kap. 6.4)	
T.36	Standardisierte Methoden und eine repräsentative Stichprobe sichern die Validität.
Durch die verwendete Standardprozedur und die hierin enthaltenen Methoden und Anleitung für die Zusammensetzung der Stichprobe kann der Evaluatoreffekt verringert werden, was für valide Ergebnisse spricht.	

## Diskussion

In der finalen Studie wurden die Testgüte der Standardprozedur und der Wahrheitsgehalt der aufgestellten Thesen überprüft. Zur Überprüfung der Testgüte wurden die Gütekriterien Objektivität, Reliabilität und Validität bestimmt.

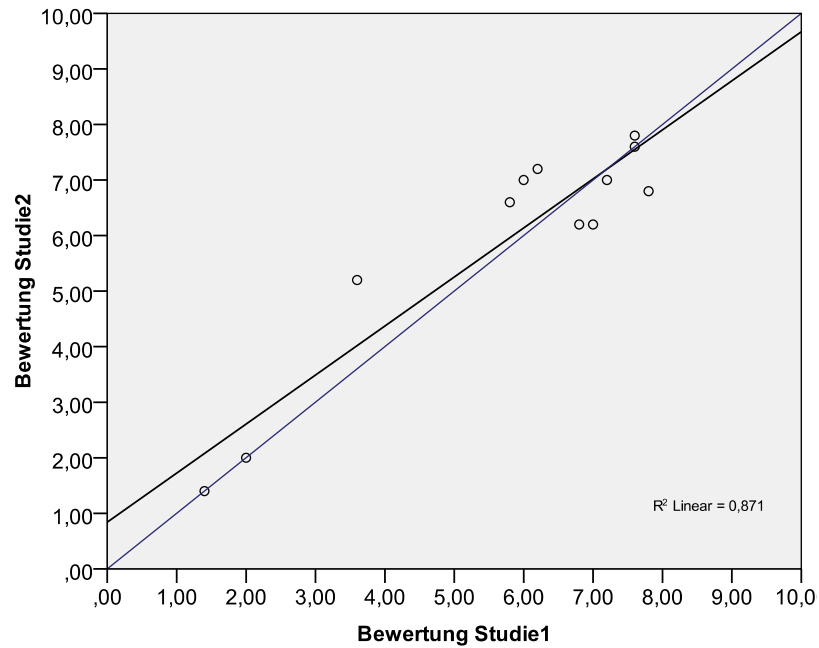
Ein Kriterium, welches die Testgüte beeinflusst, ist der Evaluatoreffekt (vgl. Kap. 4.1.3). Hertzum & Jacobsen sehen als Maßnahme zur Vermeidung dieses Effekts und der damit verbundenen Steigerung der Testgüte eine systematische Auswahl geeigneter Aufgabenszenarien. Nur durch die Wahl der relevanten Aufgaben kann sichergestellt werden, dass ähnliche Ergebnisse durch unterschiedliche Evaluatoren ermittelt werden. Durch die Vorgaben und Anleitungen der Standardprozedur wurden in beiden Studien alle Primäraufgaben des Gerätes für die Evaluation ausgewählt. Diese Übereinstimmung stellt sicher, dass in einem Nutzertest die wichtigsten Aufgaben des Gerätes evaluiert werden, und somit der Evaluatoreffekt verringert wird. Studien, in denen kein derartig standardisiertes Vorgehen angewandt wurde, zeigen hierbei weit abweichende Ergebnisse. Erkenntnisse von Molich et al. verdeutlichen, dass bei einem Vergleich von neun unterschiedlichen Evaluatoren, die das gleiche System untersuchten, nur geringe Übereinstimmungen bei den gewählten Testaufgaben auftraten. Fast die Hälfte

(49%) der gestellten Aufgaben wurden von lediglich einem Evaluator und nur 6% der Aufgaben von mehr als fünf Evaluatoren evaluiert [270]. In der in diesem Kapitel durchgeführten Studie wurden zehn der 13 insgesamt evaluierten Aufgaben durch beide Evaluatoren für einen Nutzertest herangezogen. Dies entspricht einer Übereinstimmung von 77 Prozent und sogar von 100 Prozent bei den Primäraufgaben. Für sehr komplexe Geräte könnten diese Werte jedoch anders aussehen. In diesem Fall müssten entsprechend der Standardprozedur die wichtigsten Primäraufgaben ausgewählt werden. Als Entscheidungsgrundlage für diese Wahl sollten die Erkenntnisse der Heuristischen Evaluation verwendet werden. Wenn in der Aufgabenanalyse festgestellt werden sollte, dass das zu evaluierende Gerät zu komplex ist, um in einem Nutzertest alle Primäraufgaben evaluieren zu können, so sollten die Experten der Heuristischen Evaluation die zu evaluierenden Aufgaben einschränken. Auswahlkriterien sollten potentielle Schwachstellen, Häufigkeit der Nutzung und Wichtigkeit der Aufgabe sein.

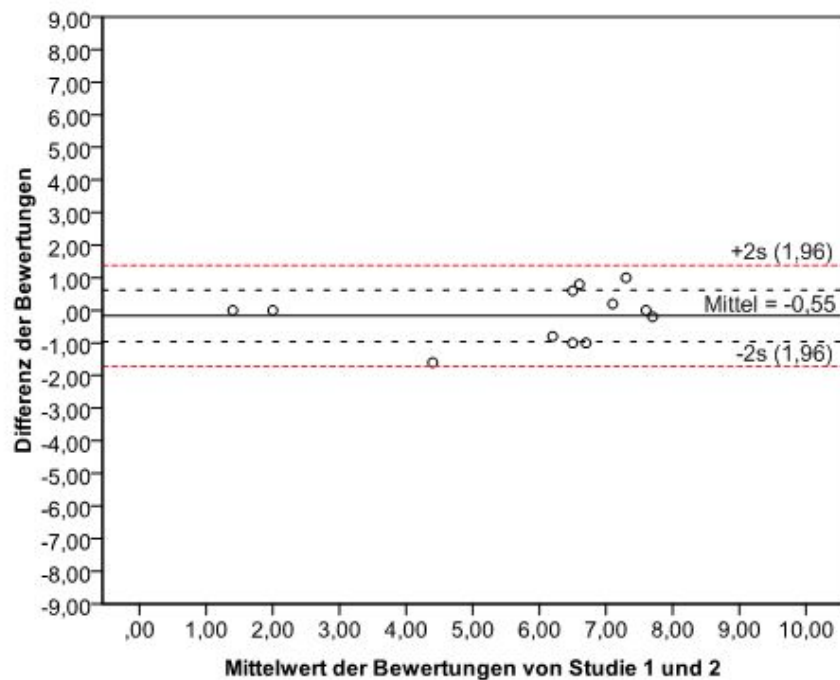
Zur Überprüfung der Objektivität wurden die Durchführungsobjektivität und die Ausführungsobjektivität betrachtet. Durch einen Vergleich der durch beide Studien gefundenen Probleme kann die Durchführungsobjektivität bestimmt werden. Insgesamt entsprach die prozentuale Übereinstimmung der gefundenen Fehler beider Studien 52%. Zusätzlich führten die Testverantwortlichen ein Ranking der wichtigsten Probleme durch. Dieses Vorgehen ist gerechtfertigt, da auch in einer Bewertung lediglich die wichtigsten Probleme behandelt werden sollten, da sonst der Bewertungsworkshop zu lange dauert, worunter die Bewertungsqualität leiden könnte. Die Erfahrung dieser Arbeit zeigt, dass bereits eine Bewertung von 15 bis 20 Problemen ca. 2 Stunden in Anspruch nimmt. Nach einem ersten Ranking durch den jeweiligen Testverantwortlichen konnte das Übereinstimmungsergebnis auf 68% bei den 20 wichtigsten und sogar auf 73% bei den 15 wichtigsten Problemen verbessert werden. In der bereits beschriebenen Studie von Molich et al. konnten lediglich 25% der insgesamt gefundenen Probleme durch zwei oder mehr Prüfer identifiziert werden. Weitere Studien zeigen ähnliche Ergebnisse [214, 268, 269, 270, 272, 182]. Dies bestätigt die Vorteile der in dieser Arbeit entwickelten standardisierten Vorgehensweise. Dennoch existierende Unterschiede sind durch unterschiedliche Bedingungen der Testreihen und durch andere Testnutzer mit anderen Präferenzen und Erfahrungen zu erklären.

Die Auswertungsobjektivität wurde dadurch bestimmt, dass die Bewertungen der beiden Studien verglichen und die Korrelation berechnet wurden, sowie durch die

Bland-Altman-Methode. Das Ziel hierbei war es, eine Korrelation von über 0,8 zu erreichen, was mit einem Ergebnis von 0,734 nicht ganz gelang, wobei dieses jedoch immerhin relativ nahe an dem gewünschten Wert lag. Auch bei der Bland-Altman-Methode wurden die Erwartungen nicht ganz getroffen. Zwar lag der Mittelwert der Messwertdifferenzen lediglich bei 0,55, jedoch war der 95%-Bereich der Messwerte mit  $\pm 3,18$  gegenüber der Erwartung zu hoch. Die meisten Probleme wurden zwar sehr ähnlich bewertet, jedoch bestimmte ein Messausreißer die Korrelation und die Bland-Altman-Methode. Die größte Abweichung gab es bei dem Problem „Spülfluss beendet bevor Katheter entfernt“ (Problem 31, vgl. Anhang S). Bei diesem Problem herrschte in den beiden Studien Uneinigkeit zwischen den Evaluatorengruppen. In der ersten Studie wurde dieses Problem von einigen Experten als sicherheitskritisch beurteilt, da sie davon ausgingen, dass der Katheter ohne Spülflüssigkeit im Patienten verweilen würde. In der wiederholten Studie gingen alle Experten davon aus, dass der Nutzer den Katheter nach Beendigung des Spülflusses sofort entfernen würde. Daher wurde dieses Problem als „kein Problem“ beurteilt. Somit liegt diese Abweichung an einer Uneinigkeit bzw. an fehlender Information der Experten und nicht an der Methode. Wird die Bewertung dieses Problems bei der Korrelationsberechnung nicht betrachtet, so liegt die Korrelation nach dieser Fehlerbereinigung bei 0,933 und ist auf dem Niveau von 0,001 signifikant. Dies entspricht einem Wert annähernd Eins. Es kann somit von einer Auswertungsobjektivität ausgegangen werden (Abb. 101). Auch in dem Bland-Altman-Plott erkennt man diese Verbesserung (Abb. 102). Der 95%-Bereich der Messwerte mit  $\pm 1,55$  um den Mittelwert der Messwertdifferenzen von 0,167 ist vollkommen akzeptabel, so dass von einer Objektivität gesprochen werden kann.



**Abb. 101: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug auf die Bewertungsergebnisse der ersten Studie (x-Achse), mit der Regressionsgeraden ( $r=0,993$ ) und der Geraden der Gleichheit (Perfekte Korrelation,  $r=1$ , blau), nach einer Bereinigung der Bewertungsergebnisse zur Objektivitätsmessung**



**Abb. 102: Bland-Altman-Methode zur Visualisierung des Zusammenhangs zweier wiederholender Studien, nach einer Bereinigung der Bewertungsergebnisse zur Objektivitätsmessung**

Zur Überprüfung der Reliabilität wurde die Retestrelabilität der Problembewertung betrachtet. Hierzu wird bei einer wiederholten Bewertung der gleichen Evaluatoren nach einem zeitlichen Abstand eine Korrelation von mindestens 0,75 angestrebt. Die Korrelation betrug in dieser Studie 0,81 und war auf dem Niveau von 0,001 signifikant. Bei diesem Wert kann von Reliabilität gesprochen werden.


Auch bei der Gesamtbewertung war eine Reliabilität feststellbar. Sowohl bei der Betrachtung aller Probleme als auch bei der alleinigen Betrachtung der nicht sicherheitskritischen Probleme lag die Gesamtwertung nahe beieinander oder war sogar gleich.

Die Validität wurde dadurch erreicht, dass die in Kapitel 4.1.8 beschriebenen Validitätskriterien eingehalten wurden. Hierzu zählten die Verwendung standardisierter, wissenschaftlich validierter Methoden im Rahmen der Standardprozedur, die Auswahl einer repräsentativen Stichprobe und eine geeignete Aufgabenauswahl für den Nutzertest. Zusätzlich wurde der Nutzungskontext in der Testsituation dargestellt. Um diesen noch realistischer zu gestalten, wurden die Aufgaben im Rahmen eines Rollenspieles verbalisiert. Zusätzlich wurde eine OP-Geräuschkulisse über einen CD-Player abgespielt, und durch den Versuchsleiter wurde Stress provoziert. Durch die Beachtung dieser Aspekte war in den Studien die Validität gewährleistet.






Insgesamt erwies sich die Standardprozedur in der finalen Studie als ein geeignetes Instrument, um die Gebrauchstauglichkeit medizinischer Geräte unter gleichzeitiger Betrachtung der Sicherheit zu evaluieren und zu bewerten.

### **Verifikation der Anforderungen**







Als letzter Schritt der Validierung wurden die in Kapitel 3.3 und Kapitel 6.2 aufgestellten Anforderungen an die Standardprozedur überprüft. Hierbei dienten die Ergebnisse der einzelnen Studien dazu, diese zu be- oder zu widerlegen.






<b>A1. Finden ausreichend vieler Probleme</b>	
Um diese Anforderung zu erfüllen, muss die Standardprozedur die meisten und schwerwiegendsten Probleme der Hauptaufgaben und der wichtigsten Nebenaufgaben aufdecken. Dies wird durch die passende Wahl der Arbeitsaufgaben für einen Nutzertest erreicht (vgl. Kap. 4.1.1.2 u. 4.1.2). Hierzu werden die Primäraufgaben, die wichtigsten Sekundäraufgaben und die potentiell problembehafteten Aufgaben, welche in einer Heuristischen Evaluation (vgl. Kap. 4.1.2) identifiziert wurden, in einem Nutzertest überprüft. Entsprechend einer Berechnungsmethode	









<p>nach Lewis [241] kann gezeigt werden, dass mit 10 Testnutzern über 80% der Probleme bei den gewählten Aufgaben entdeckt werden (vgl. Kap. 4.2 u. 6.4.) [vgl. Thesen 24, 25, 29].</p>	
<p><b>A2. Durchführbarkeit im tatsächlichen Nutzungskontext</b> Für die Standardprozedur werden repräsentative Arbeitsaufgaben und Anwender ausgewählt (vgl. Kap. 4.1.1.1 u. 4.1.1.2). Auch wird versucht, die tatsächliche Nutzungsumgebung (vgl. Kap. 4.1.1.3) in einer Laboreinrichtung zu simulieren. Eine Laborumgebung wird allerdings niemals den tatsächlichen Nutzungskontext wiedergeben können. Dennoch wird dieses Defizit akzeptiert, da die Alternative eines Tests im realen Nutzungskontext aus ethischen und sicherheitsbezogenen Gründen nicht möglich ist [vgl. These 30].</p>	
<p><b>A3. Gebrauchstauglichkeit der Standardprozedur</b> Die Standardprozedur unterstützt Evaluatoren bei der Ermittlung der Gebrauchstauglichkeit medizinischer Geräte. Hierzu werden in dieser Arbeit die Methoden genau beschrieben und hilfreiche Beispiele und Hinweise gegeben. In Kapitel 6.2.2 beurteilen 32 Anwender das Bewertungsverfahren als einfach anwendbar und unterstützend bei der Bewertung der Gebrauchstauglichkeit medizinischer Geräte.</p>	
<p><b>A4. Ethische Vertretbarkeit und Vermeidung von Störungen im Arbeitsablauf des Krankenhauses</b> Durch die Testdurchführung in einem Labor und die Simulation von Behandlungen wird der Arbeitsablauf im medizinischen Arbeitssystem nicht gestört und Ethik und Sicherheit werden gewährleistet. Daten der Testnutzer werden nicht veröffentlicht und Dritten nicht zugänglich gemacht.</p>	
<p><b>A5. Angemessenheit bzgl. Kosten-Nutzen-Relation innerhalb der gegebenen Fragestellung</b> Der Aufwand wird durch die Einfachheit und Schnelligkeit der Datenerfassung und Verarbeitung sowie durch die Standardisierung der Erfassungs- und Bewertungsmethode reduziert [358]. Die Standardprozedur ist zwar durch die Verknüpfung zweier Evaluationsmethoden etwas aufwändiger als die meisten Gebrauchstauglichkeitsmethoden zur summativen Evaluation (vgl. Kap. 4.1.2), jedoch ist dieser Aufwand durch den hierdurch entstehenden Nutzen gerechtfertigt, angesichts der Tatsache, dass es sich um sicherheitskritische Geräte handelt.</p>	
<p><b>A6. Erwerb einer hohen Akzeptanz</b> Die Entwicklung und Evaluation der Standardprozedur wurde unabhängig von industriellen Interessen am Experimental-OP der medizinischen Fakultät des Universitätsklinikums Tübingen durchgeführt. Durch die industrielle Unabhängigkeit und die in dem universitären Umfeld vorhandenen fachlichen Fähigkeiten kann von einer hohen Glaubwürdigkeit und Kompetenz der Evaluatoren ausgegangen werden. Weiterhin liegt der Entwicklung ein geplantes und strukturiertes Vorgehen zugrunde, welches im Rahmen dieser Arbeit veröffentlicht wird. Zusätzlich wird die positive Einstellung unterschiedlicher Interessensgruppen in Kapitel 6.2.2 gezeigt.</p>	



<p>Außerdem hat sich die Standardprozedur in verschiedenen Studien im Rahmen dieser Arbeit bewährt. Daher ist eine hohe Akzeptanz zu erwarten.</p>	
<p><b>A7. Objektivität, Reliabilität und Validität</b>  Standardisierte Elemente, die Auswahl einer repräsentativen Stichprobe und geeigneter Aufgaben, die Beachtung des Nutzungskontextes und die Wahl geeigneter Methoden verbessern die Testgüte. Diese Punkte werden von der Standardprozedur erfüllt. Dass hierbei ein weitestgehend zufriedenstellendes Maß erreicht wurde, konnte in diesem Kapitel gezeigt werden [Thesen 32-36].</p>	
<p><b>A8. Nützlichkeit</b>  Die in dieser Arbeit entwickelte Methode ermöglicht es, standardisiert die Gebrauchstauglichkeit medizinischer Geräte zu evaluieren und zu bewerten. Die Ergebnisse der Methode können von Herstellern nicht nur zu Marketingzwecken, sondern auch zu Produktverbesserungen eingesetzt werden. Anwender und Betreiber können die Ergebnisse der Methode in ihre Kaufentscheidung einbeziehen, und Zertifizierer können sie für die Zertifizierung nach der DIN EN IEC 60601-1-6 und der DIN EN IEC 62366 verwenden.</p>	
<p><b>A9. Fairness und Neutralität (Unparteiische Durchführung und Berichterstattung)</b>  Die Bewertung führen mehrere unabhängige Experten durch. Ausreißer fallen durch die Mittelwertbildung nicht zu stark in das Gewicht der Bewertung (vgl. Kap. 6.2). Die Testgütekriterien (vgl. Kap. 6.4) werden eingehalten.</p>	
<p><b>A10. Offenlegung der Ergebnisse und Methoden (Transparenz)</b>  Die Standardprozedur wurde im Rahmen dieser Dissertation entwickelt und evaluiert. Die Methode, die Entwicklungsschritte und die Ergebnisse der Evaluation sind in dieser Dissertationsschrift veröffentlicht.</p>	
<p><b>A11. Berücksichtigung des Standes von Forschung und Technik</b>  Der Entwicklung der Standardprozedur ging eine umfangreiche Literaturrecherche voraus. Aktuelle Methoden des Usability Engineerings und der Risikobetrachtung werden verglichen und diskutiert. Weiterhin berücksichtigt die Standardprozedur die Anforderungen der aktuellen internationalen Standards DIN EN IEC 60601-1-6 und DIN EN IEC 62366.</p>	
<p><b>A12. Einsetzbarkeit für eine Vielzahl medizinischer Geräte</b>  Die Standardprozedur gilt für eine Vielzahl von Medizinprodukten. Lediglich durch eine Expertenbegutachtung wird es nicht möglich sein, die Komplexität jedes Medizinproduktes zu erfassen. Insbesondere wäre es unmöglich, sich in die unterschiedlichen Nutzer und Umgebungen des heterogenen medizinischen Arbeitssystems hineinzusetzen. Nur durch die Einbindung der jeweiligen Anwender des Produktes in einer simulierten Arbeitsumgebung ist dies möglich. Daher kombiniert die Standardprozedur die Methoden Heuristische Evaluation und Nutzertest miteinander. In der Studie zur Gebrauchstauglichkeitsevaluation der minimalinvasiven Handgriffe wird gezeigt, dass die Standardprozedur auch für passive Medizinpro-</p>	

<p>dukte anwendbar ist (vgl. Kap. 4.2.2).</p>	
<p><b>A13. Abdeckung des versehentlichen Missbrauchs</b>                  Die DIN EN IEC 60601-1-6 fordert eine Betrachtung des vernünftigerweise vorhersehbaren Missbrauchs [97]. Dieser ist jedoch in den meisten Fällen in einem Gebrauchstauglichkeitstest schwer zu erfassen. Daher kann er lediglich in die Expertenevaluation mit einbezogen werden.</p>	
<p><b>A14 Berücksichtigung der Anforderungen an ein sicherheitskritisches System</b>                  In dieser Arbeit werden medizinische Geräte als sicherheitskritische Geräte eingestuft. Sicherheit ist die Freiheit von Risiken. Die Gebrauchstauglichkeitsbewertung berücksichtigt die Risikobeurteilung und stuft Gebrauchstauglichkeitsprobleme durch die Faktoren Problemschwere, Auftretenshäufigkeit und mögliches Schadensausmaß ein.</p>	
<p><b>A15 Bewertung der Gebrauchstauglichkeit</b>                  Die Bewertung der Gebrauchstauglichkeit orientiert sich an den, in den internationalen Normen DIN EN IEC 62366, DIN EN IEC 60601-1-6 und DIN EN 9241-11 festgelegten Merkmalen für Gebrauchstauglichkeit. Sowohl die Auswahl der Bewertungsmerkmale als auch die Quantifizierung stellen subjektive Verfahren dar. So ist nicht von der Hand zu weisen, dass sich in jedem Bewertungsverfahren Erfahrungen, Wertmaßstäbe und Urteile der Experten niederschlagen. Dieser Sachverhalt ist nach Schmidtke akzeptabel, da diesem Bewertungsverfahren eine immanente Logik zugrunde liegt, und die Kennzeichnung der Einzelmerkmale und deren Stufung so eindeutig erfolgt, dass Experten in der konkreten Anwendung zu hinreichend vergleichbaren Ergebnissen kommen [358], was auch die Evaluation der Testgütekriterien zeigt (vgl. Kap. 6.4).                  Alle von 32 befragten Experten beurteilte das hier entwickelte Bewertungsverfahren als nützlich, um die Gebrauchstauglichkeit medizinischer Geräte zu bewerten (vgl. Kap. 6.2.2).</p>	
<p><b>A16 Bewertung nach Normkonformität mit der DIN EN IEC 60601-1-6 und DIN EN IEC 62366</b>                  Die Bewertungsmethode wird unter Berücksichtigung der Anforderungen der DIN EN IEC 60601-1-6 und der DIN EN IEC 62366 entwickelt. Durch die Betrachtung der Sicherheit, bezogen auf die Gebrauchstauglichkeit medizinischer Geräte, kann die Normkonformität mit diesen Normen bestätigt werden. Dieser Bereich wird in der Bewertung der UseProb hervorgehoben.</p>	
<p><b>A17 Allgemeinverständliches Bewertungssystem</b>                  Alle der 32 befragten Experten beurteilen dieses Tool als einfach anzuwenden und verständlich (vgl. Kap. 6.2.2). Auch in den weiteren durchgeführten Entwicklungsstudien (vgl. Kap. 6.2 u. Kap. 6.4) gibt es keine Probleme. Die Bewertung wird durch das Bewertungstool der UseProb unterstützt. In diesem Tool werden alle</p>	

<p>wichtigen Begriffe erklärt und definiert. Vor der eigentlichen Bewertung sollte eine Beispielbewertung durchgeführt werden, um sicherzustellen, dass alle Experten die Methode verstehen. Die Ergebnisse der Gebrauchstauglichkeitsbewertung werden in einem 10-stufigen Notensystem ausgegeben, welches zusätzlich zur Förderung der Interpretationsfähigkeit mit farbigen Markierungen versehen ist.</p>	
<p><b>A18 Verknüpfung von Gebrauchstauglichkeits- und Risikobewertung</b></p> <p>Durch die Verbindung der Gebrauchstauglichkeitsbewertung (vgl. Kap. 5.8) und der Risikobewertung (vgl. Kap. 5.4.2) und durch die Berücksichtigung der in den Normen DIN EN IEC 60601-1-6, DIN EN IEC 62366 und DIN EN 14971 festgelegten Anforderungen wird eine Bewertung der Gebrauchstauglichkeit medizinischer Geräte unter Beachtung der Sicherheit für Patienten, Anwender und Material durchgeführt.</p>	
<p><b>A19 Bezug zu den entsprechenden Anforderungen der relevanten Normen für Sicherheit und Gebrauchstauglichkeit</b></p> <p>Die Anforderungen der allgemeinen Normen für Gebrauchstauglichkeit [105, 113, 114, 200], der speziellen Normen für die Gebrauchstauglichkeit medizinischer Geräte [97, 104] und der Norm für das Risikomanagement für Medizinprodukte [108] werden bei der Gebrauchstauglichkeitsbewertung berücksichtigt.</p>	
<p><b>A20 Aufbau der Bewertung auf den Ergebnissen der Standardprozedur</b></p> <p>Die Methode bewertet die Gebrauchstauglichkeit medizinischer Geräte aus den Ergebnissen der Standardprozedur anhand der Gebrauchstauglichkeitsmaße Effektivität, Effizienz und Zufriedenheit unter Berücksichtigung der Sicherheit. Es wird gezeigt, dass mit Hilfe der Standardprozedur die meisten Gebrauchstauglichkeitsprobleme eines medizinischen Gerätes entdeckt werden können (vgl. Kap. 4.2), und dass der Evaluatoreffekt verringert wird (vgl. Kap. 6.4).</p>	
<p>Legende:</p> <p>  Voll erfüllt       teilweise erfüllt       nicht erfüllt </p>	

Die Verifikation dieser Anforderungen hat gezeigt, dass durch die Standardprozedur fast alle aufgestellten Anforderungen in vollstem Umfang erfüllt werden. Ausnahmen sind die Anforderungen A2, A7 und A13. Anforderung A2, die Evaluation im Rahmen des tatsächlichen Nutzungskontextes durchzuführen, ist nicht zu erfüllen, da dies aus sicherheitskritischen und ethischen Aspekten im Rahmen der Patientengefährdung nicht zu verantworten ist. Die Erfüllung der Testgütekriterien (Anforderung A7) wird auch nicht ganz erreicht. Bei der Ermittlung der Auswer-

tungsobjektivität liegt der erreichte Wert der Korrelation gering unter dem geforderten Wert. Dies ist jedoch eher auf eine Unstimmigkeit innerhalb der Experten-Gruppe zurückzuführen (vgl. Kap. 6.4). Eine solche Fehlerquelle ist nie gänzlich auszuschließen, daher sollten bei einer vermuteten Unstimmigkeit zusätzliche Experten zur Bewertung hinzugezogen werden. Ansonsten werden bei der Ermittlung der Testgütekriterien die Anforderungen in vollem Maße erfüllt. Die Abdeckung des versehentlichen Missbrauchs (Anforderung A13) kann die Standardprozedur nicht gänzlich erfüllen. Diese Betrachtungen sind lediglich innerhalb der Heuristischen Evaluation möglich. Weitere Untersuchungen sollten im Rahmen des Risikomanagements erfolgen.

## 7 Abschließende Diskussion und Fazit

Viele Elemente dieser Arbeit wurden bereits in vorherigen Kapiteln im Rahmen des iterativen Entwicklungsprozesses diskutiert. Dies war notwendig, um Entscheidungen für die Weiterentwicklung der Standardprozedur zu treffen und neue Ansätze herzuleiten. In dieser abschließenden Diskussion wird daher nur auf die wesentlichen Aspekte dieser Arbeit eingegangen. Für spezifische Informationen sei auf die Diskussionen innerhalb dieser Arbeit verwiesen.

Die Notwendigkeit dieser Arbeit begründet sich primär durch die in Kapitel 2.2.2.2 gezeigten Gefahren durch Gebrauchstauglichkeitsprobleme in der Medizin. Verschiedene Studien zeigen, dass gerade im medizinischen Umfeld ein erhöhtes Defizit bei der Berücksichtigung von Gebrauchstauglichkeitsanforderungen herrscht [136, 158, 226, 227, 246, 340]. Geißler et al. stellen fest, dass die Einflussfaktoren von Fehlern aufgrund von Gebrauchstauglichkeitsproblemen in der Medizin nur wenig analysiert worden sind, und daher eine Risikoabschätzung kaum möglich ist [149]. Gebrauchstauglichkeit und ergonomische Aspekte müssen prospektiv überprüft und analysiert werden. Das Problem hierbei ist, dass in der Entwicklungsphase keine Feldinformationen zu Gebrauchstauglichkeitsproblemen, Fehlerwahrscheinlichkeit und Einflussfaktoren im Umfeld der Anwendung zur Verfügung stehen [202]. Hallbeck et al. fordern ein Bewertungssystem, welches auf Grundlage der Normen DIN EN IEC 60601-1-6 und DIN EN IEC 62366 die Evaluation und Bewertung der Gebrauchstauglichkeit medizinischer Geräte standardisiert und als Grundlage für Zertifizierer dienen kann [14]. Die Ergebnisse dieser Arbeit können dazu verwendet werden, dieses Defizit zu beheben. Bei konsequenter Umsetzung sind sie ein Element der vom IOM geforderten Standardisierungen zur Vermeidung unerwünschter Ereignisse bei Medizinprodukten und damit zur Erhöhung der Patientensicherheit [10].

Das Ziel dieser Arbeit ist es, eine allgemeingültige Standardprozedur zur Überprüfung der Gebrauchstauglichkeit medizintechnischer Produkte zu entwickeln. Zur Bewertung der Gebrauchstauglichkeit wird ein zweigeteilter Methodenansatz aus der Problemevaluation und der Problembewertung gewählt. Mit dieser Methode soll im Sinne einer summativen Evaluation die Erfüllung der Anforderungen an die Gebrauchstauglichkeit im Rahmen der Zweckbestimmung eines medizinischen Gerätes validiert werden. Hierzu werden Methoden und Erkenntnisse des Risikomanagements mit denen der Gebrauchstauglichkeitsbetrachtung verknüpft.

Bei der Entwicklung der Standardprozedur werden zuerst die theoretischen Rahmenbedingungen betrachtet. Danach werden die beiden Phasen der Standardprozedur, nämlich die Problemevaluation und die Problembewertung getrennt voneinander entwickelt und evaluiert. Hierzu wird ein, an den bewährten benutzerzentrierten Designprozess (vgl. Kap. 2.1.4) angegliederter, iterativer Entwicklungsprozess verwendet. In wiederkehrenden Entwicklungs- und Evaluationsphasen, unter Einbeziehung der Benutzer, werden die beiden Phasen konkretisiert. Am Ende dieser Arbeit wird die gesamte Standardprozedur validiert. Dieser Prozess eignet sich für die Entwicklung durch seine strukturierte und schrittweise evaluierende Vorgehensweise. Durch die Einbeziehung von Experten und Nutzern und durch die Auswahl geeigneter Test- und Evaluationsmethoden wird ein Verfahren entwickelt, welches die gesetzten Anforderungen erfüllt (vgl. Kap. 3.3 u. 6.4). Diese Methode legt ein standardisiertes Verfahren fest, durch welches die Elemente einer Evaluation obligatorisch werden, was auch zur Reliabilität der Evaluation beiträgt und die Ergebnisse mit anderen Evaluationen vergleichbar macht.

Für die Evaluationsphase der Standardprozedur werden bewährte und standardisierte Methoden des Usability Engineerings geprüft und für den speziellen Zweck an die Anforderungen des medizinischen Arbeitssystems adaptiert. Das Grundgerüst der Evaluation ist die Erfassung des Nutzungskontextes. Diese für die Gebrauchstauglichkeitsevaluation unerlässliche Analyse und Beschreibung der Benutzer, der Arbeitsaufgaben, der Arbeitsmittel und der Arbeitsumgebung liefern elementare Informationen für die Überprüfung der Gebrauchstauglichkeit [33, 97, 113, 105, 202, 248, 346, 426, 430]. Erst hierdurch ist eine Festlegung der Evaluationsziele, der Aufgaben- und Testszenarien sowie die Wahl der Testumgebung und der Testnutzer möglich.

Für die Evaluation der Gebrauchstauglichkeit, entsprechend der in den internationalen Normen festgelegten Bewertungskriterien Effektivität, Effizienz und Zufriedenheit [97, 104, 113], werden die Methoden Heuristische Evaluation und Nutzer-test gewählt (vgl. Kap. 2.4.1). Verschiedene Studien zeigen, dass durch die Kombination dieser Methoden die meisten Gebrauchstauglichkeitsprobleme gefunden werden [12, 120, 124, 142, 184, 270, 372, 284]. Gebrauchstauglichkeitsevaluationen lassen sich schon häufig für medizinische Geräte finden. Des Öfteren werden dabei auch die in dieser Arbeit angewendeten Methoden Heuristische Evaluation [11, 158, 389, 426, 428] und Nutzertest [11, 188, 226, 266, 340, 402, 413, 426]

eingesetzt. Auch wird in derartigen Studien eine Bewertung der Gebrauchstauglichkeit durchgeführt [11, 158, 426], jedoch findet hierbei lediglich eine Bewertung statt, wie man sie auch für Konsumgeräte anwendet (vgl. Kap. 5.8). Sicherheitskritische Aspekte werden dabei nicht beachtet. Diese werden erst im Rahmen des Risikomanagements berücksichtigt (vgl. Kap. 5.5). Häufig werden in solchen Fällen die Mensch-Maschine-Kommunikation und die Gebrauchstauglichkeit fälschlicherweise ignoriert. Gebrauchstauglichkeit ist jedoch ein wichtiges Fundament für die Sicherheit in der medizinischen Arbeitsumgebung. Erst die Standardprozedur betrachtet sie unter diesem Aspekt.

Das Problem bei der Heuristischen Evaluation ist, dass häufig die hier von Experten gefundenen Probleme nicht in einem Nutzertest beobachtet werden [72, 73, 74, 210, 285]. Daher dient sie in der Standardprozedur lediglich dazu, potentielle Schwachstellen zu ermitteln, welche mit einem Nutzertest verifiziert werden müssen. Nur durch den Nutzertest ist es möglich, diese zu bewerten, da hier Informationen über die Tragweite und die Auswirkungen der Probleme erhalten werden können. Als positiv stellt sich bei der Heuristischen Evaluation heraus, dass ein System durch ein strukturiertes Vorgehen untersucht wird. Freie Expertenevaluationen sind dagegen nicht so ergiebig (vgl. Kap. 2.4.1.1). Es empfiehlt sich allerdings, speziell auf den medizinischen Nutzungskontext angepasste Heuristiken zu verwenden, da die bekannten Heuristiken zu interaktiven Systemen [62, 271, 282, 284, 311, 319, 368, 392] für medizinische Experten ungewohnt sind (vgl. Kap. 4.2.1).

Die beschriebenen Methoden des Nutzertests stellen sich in der vorliegenden Arbeit als geeignet heraus, um die Gebrauchstauglichkeit medizinischer Geräte zu evaluieren (vgl. Kap. 2.4.1.2). Nutzertests bieten im Gegensatz zu dem formal-analytischen Verfahren der Heuristischen Evaluation den Vorteil, dass reale Nutzer reale Aufgaben durchführen, und dass somit Rückschlüsse auf Problemschwere und Auftretenshäufigkeit gewonnen werden können.

In den in Kapitel 4.2 beschriebenen Studien kann gezeigt werden, dass sich mit dem Nutzertest eine hinreichend große Menge an Gebrauchstauglichkeitsproblemen evaluieren lassen. In den Entwicklungsstudien (vgl. Kap. 6.2) wird gezeigt, dass mit diesen Daten eine Gebrauchstauglichkeitsbewertung möglich ist. Ein in der Literatur häufig beschriebenes Problem beim Nutzertest ist der Evaluatoreffekt [182, 214, 268, 269, 270, 272]. Dieser Problematik wird bei der Entwicklung der Standardprozedur besondere Beachtung gewidmet. Durch die standardisierte

Vorgehensweise (vgl. Kap. 4.3), eine genaue Zieldefinition (vgl. Kap. 4.1.2), die definierte Wahl der Aufgabenszenarien (vgl. Kap. 4.1.5), eine realgetreue Versuchsumgebung (vgl. Kap. 4.1.6), die Wahl repräsentativer Testnutzer (vgl. Kap. 4.1.4) und Bewertungsexperten (vgl. Kap. 4.1.3) sowie durch eindeutig definierte Problemkriterien (vgl. Kap. 5.2 u. 5.1.4) wird diesem Effekt entgegengewirkt und die Ergebnisqualität gesteigert. Als Maß für den Evaluatoreffekt gelten Objektivität und Reliabilität der Gebrauchstauglichkeitsuntersuchung [181, 182]. Bei der Fehleridentifikation ist jedoch die Erfassung der Reliabilität nicht möglich. Hierfür müsste ein wiederholter Test unter gleichen Bedingungen die gleichen Ergebnisse reproduzieren können [36]. Diese gleichen Bedingungen sind jedoch bei einem Gebrauchstauglichkeitstest nicht möglich, da die Stichprobe nach einem ersten Test bereits die Testprozedur und mögliche Probleme kennt und in einem wiederholten Test andere Verhaltensweisen zeigen würde. Die Erreichung der Objektivität wird in der Praxis dadurch angezeigt, dass bei verschiedenen Prüfern Deckungsgleichheit der evaluierten Fehler eintritt [214, 268, 269, 270, 272, 182]. In Kapitel 6.4 wird gezeigt, dass mit Hilfe der Standardprozedur hierbei ausreichende Deckungsgleichheit gegeben ist.

Die Evaluationsphase identifiziert Probleme bei der Geräteverwendung und gibt Hinweise zu deren Schweregrad [372], jedoch kann kein spezieller Bezug zu einer damit verbundenen möglichen Patientengefährdung hergestellt werden. Im Anschluss an die Evaluation der Gebrauchstauglichkeit erfolgt daher im Sinne der Standardprozedur die Bewertung der Gebrauchstauglichkeit. Bisher werden in den meisten Gebrauchstauglichkeitsstudien in der Medizin Probleme gar nicht oder lediglich nach den bekannten Schemata des Usability Engineerings bewertet, welche keine Risikobetrachtung berücksichtigen [11, 158, 426]. In der Medizin ist die Verwendung von Geräten eng mit der Sicherheit verknüpft. Daher wird in dieser Arbeit die Gebrauchstauglichkeitsbetrachtung mit der Risikobetrachtung verbunden, um Gebrauchstauglichkeitsdefizite zu bewerten (vgl. Kap. 6). Bisherige Bewertungsmethoden des Usability Engineering sind weniger auf sicherheitskritische Geräte ausgerichtet und lassen den Aspekt der Sicherheit vermissen [72, 124, 183, 184, 284, 348]. Methoden des Risikomanagements betrachten Gebrauchstauglichkeit, wenn auch nur im geringen Maße [35, 101, 102, 103, 107, 234, 267]. Bisher existiert keine Methode, welche eine Bewertung entsprechend den internationalen Normen DIN EN IEC 60601-1-6 und DIN EN IEC 62366 für die Gebrauchstauglichkeit medizinischer Geräte unter dem Aspekt der Sicherheit



ermöglicht. Die Standardprozedur kombiniert hierzu bewährte Methoden des Risikomanagements (Kap. 5.4) mit denen des Usability Engineerings (Kap. 5.8). Ebenso wie der Evaluationsteil der Standardprozedur wird auch der Bewertungsteil in einem iterativen Prozess entwickelt. Das Ergebnis ist die UseProb-Methode, welche, aufbauend auf den Ergebnissen des Evaluationsteils, eine Bewertung der Gebrauchstauglichkeit medizinischer Geräte ermöglicht.

Die Methode wird durch verschiedene Studien evaluiert. In der ersten Entwicklungsstudie wird sie durch eine Konsensbildung aus verschiedenen Methoden ausgewählt (vgl. Kap. 6.2.1). In der zweiten Entwicklungsstudie wird gezeigt, dass mit dieser Methode eine Bewertung von Gebrauchstauglichkeitsproblemen medizinischer Geräte möglich ist. Weiterhin beurteilen 32 Experten aus der Medizin und der Mensch-Maschine-Kommunikation die Methode durchweg als unterstützend und nützlich, um die Gebrauchstauglichkeit medizinischer Geräte zu bewerten (vgl. Kap. 6.2.2). Auch für diese Methode werden die Testgütekriterien Objektivität, Validität und Reliabilität betrachtet (vgl. Kap. 6.4). Als Maß für Reliabilität und Objektivität gilt die Korrelation zweier Messwertreihen. Die Reliabilität wird mit dem Korrelationskoeffizienten ( $r$ ) gemessen (vgl. Kap. 4.1.8). Um von einem reliablen Test sprechen zu können, sollte eine Reliabilität von  $r > 0,75$  erreicht werden. Im Gegensatz zum Evaluationsteil der Standardprozedur ist bei dem Bewertungsteil eine Reliabilitätsmessung mit Hilfe der Korrelation durch einen wiederholten Test möglich, da die Testergebnisse unabhängig von den Erfahrungen der Experten mit dem Produkt sein sollten. In einer wiederholten Bewertung mit derselben Expertengruppe wird auf diese Weise eine Korrelation von  $r = 0,801$  erreicht, sodass von einem reliablen Ergebnis der beiden Studien gesprochen werden kann. Die Objektivitätsmessung erfolgt durch eine wiederholte Bewertung mit einem anderen Evaluator und anderen Bewertungsexperten. Hierbei liegt die Korrelation bei 0,734 und nach einer Fehlerbereinigung bei 0,933. Grafisch wird die Testgüte durch die Bland-Altman-Methode veranschaulicht (vgl. Kap. 6.4).

Durch diese Berechnungen und die grafische Veranschaulichung kann gezeigt werden, dass die Methode die Testgütekriterien einhält und den Evaluatoreffekt verringert. Dennoch bleibt diese Bewertung eine subjektive Handlung, die niemals die Präzision eines Metermaßes oder einer Stoppuhr erreichen kann, da sie von den Einstellungen, Erfahrungen und Wertmaßstäben der Bewertenden abhängt. Somit sind Abweichungen nicht gänzlich auszuschließen. Laut Schmidte ist dieser Sachverhalt solange akzeptabel, als dem Bewertungsverfahren eine im-

manente Logik zugrunde liegt, und die Kennzeichnung der Einzelmerkmale und deren Stufung so eindeutig erfolgt, dass Fachleute in der konkreten Anwendung zu hinreichend vergleichbaren Ergebnissen kommen [358]. Die Logik des Bewertungsverfahrens wird in dieser Arbeit dargestellt, und die Vergleichbarkeit der Ergebnisse kann durch Bestimmung der Testgütekriterien gezeigt werden. Diese Studien bieten jedoch nur erste Anhaltspunkte. Um genauere Aussagen zu treffen, sind weitere Studien notwendig. Harloff schlägt für eine eindeutige Bestimmung der Testgüte fünf Studien vor, in denen fünf unterschiedliche Evaluatoren die Gebrauchstauglichkeit jeweils eines Gerätes evaluieren [168]. Neben dem Mangel an geeigneten Prüfinstitutionen in Deutschland schränkt auch der enorme monetäre Aufwand einer derartigen Untersuchung die Möglichkeiten im Rahmen dieser Arbeit deutlich ein. Es bleibt somit abzuwarten, wie sich die Standardprozedur im praktischen Einsatz bewährt. Abweichungen zwischen unterschiedlichen Prüfinstituten wird es immer wieder geben. Um diese Unterschiede zu nivellieren, soll die Standardprozedur eine pragmatische Methode sein, um eine standardisierte Evaluation und eine Bewertungsbasis zu ermöglichen.

In der Praxis hat sich die Methode bereits bewährt. Zur gleichen Zeit, in der mit der Standardprozedur die Spritzenpumpe evaluiert wurde (vgl. Kap. 4.2.3), untersuchte auch die Bundesanstalt für Arbeitsschutz und Arbeitsmedizin das gleiche Produkt mit einem Nutzertest [413]. Es ist festzustellen, dass mit dem in dieser Arbeit entwickelten und standardisierten Vorgehen für eine bewertende Gebrauchstauglichkeitsevaluation mehr Probleme gefunden wurden als in der Studie der Bundesanstalt [45]. Die Bewertung konnte zeigen, dass es sich hierbei auch um sicherheitskritische Probleme handelte. Diese Abweichung kam u.a. auch dadurch zustande, dass in der Studie der Bundesanstalt nicht alle Primäraufgaben untersucht wurden. So wurde beispielsweise keine Bolusgabe untersucht, obwohl sie eine Primäraufgabe darstellt. Daher ist es nicht verwunderlich, dass die hier möglichen Probleme nicht alle erkannt wurden. Auch wurde keine Bewertung der Probleme durchgeführt, sodass deren Tragweite auch nicht bestimmt werden konnte.

Obwohl die Standardprozedur die Evaluation und Bewertung medizinischer Geräte weitestgehend standardisiert, sind einige Messfehler (Bias) möglich, deren Berücksichtigung die Qualität der Ergebnisse verbessert. Eine besondere Rolle bei der Evaluation spielt der Testverantwortliche. Durch ihn können verschiedene Messfehler entstehen. Während eines Nutzertests kann er z.B. durch zu frühe

Hilfe oder ungenaue Zielvorgaben den Test verfälschen. Andererseits können während der Bewertung der Gebrauchstauglichkeitsprobleme die Experten zu sehr auf den Testverantwortlichen fokussiert sein, und dieser kann durch Suggestion die Meinung der Teilnehmer beeinflussen. Daher sollte der Testverantwortliche ein neutrales Verhalten einnehmen und darauf achten, dass der Test nicht verfälscht wird. In dieser Arbeit werden hierzu an verschiedenen Stellen Verhaltenshinweise gegeben.

Ein Negativaspekt der Gebrauchstauglichkeitsevaluation ist die Tatsache, dass der Nutzungskontext lediglich nachgestellt und kein wirklicher Patient behandelt wird. Janß & Lauer bemängeln, dass realitätsnahe, kontextspezifische Stress- oder auch Notfallsituationen in einer experimentellen Laborumgebung meist nur unzureichend dargestellt werden können [202]. Im Rahmen dieser Arbeit wurde mitunter festgestellt, dass einige Probanden sehr sorglos mit den Geräten umgingen, so dass es bei einer Realanwendung zu einer gefährlichen Situation für den Patienten gekommen wäre. Wahrscheinlich hätten die Probanden sich in der realen Nutzungssituation anders verhalten, wenn ein wirklicher Patient gefährdet gewesen wäre. Daher kann man nicht immer das Verhalten in der Testsituation auf ein Verhalten in der realen Nutzungssituation beziehen. Dennoch können hier Fehlerursachen und Probleme entdeckt werden. Eine in einer Labortestsituation falsch durchgeführte Aufgabe wird vielleicht in der realen Testsituation gar nicht erst durchgeführt. Die Problematik ist hierbei gleich, lediglich die Handlungen der Anwender sind unterschiedlich. Auch ist die Möglichkeit, eine Geräteevaluation in der realen Nutzungssituation durchzuführen, aus ethischen Gründen keine Alternative zu der Laboruntersuchung, da dies wegen der möglichen Patientengefährdung nicht vertretbar ist (vgl. Kap. 6.4 Anforderung A4). Bei der Laboruntersuchung kann dieser Effekt dadurch vermindert werden, dass eine möglichst realistische Nutzerumgebung geschaffen wird. Hierzu gehören beispielsweise Patientendummies, eine realistische Arbeitsumgebung, ein realistischer Geräuschpegel und die Verbalisierung einer Aufgabenstellung in Form eines Rollenspiels statt in papierbasierter Form (vgl. Kap. 6.4). Letzterer Punkt zeigte hierbei wesentliche Verbesserungen. Durch ein Rollenspiel schienen die Probanden mehr in der Nutzungssituation integriert und der Effekt war nicht mehr zu beobachten.

Weiterhin bemerken Kaye und Crowley, dass die durch die Nutzerstudien ermittelten Ergebnisse zur Auftretenswahrscheinlichkeit von Fehlern irreführend sein können [212]. Gerade bei neuen Medizingeräten liegen noch keine Daten über die

Verwendung von erfahrenen Nutzern vor. Dieses Defizit wird dadurch vermindert, dass in die Bewertung die Erfahrung verschiedener Fachexperten einfließt, welche die Probleme kritisch betrachten und solche Einschränkungen mit berücksichtigen.

Auch kann es vorkommen, dass bei der Gebrauchstauglichkeitsbewertung zufälligerweise eine Gruppe zusammengesetzt wird, welche nicht repräsentativ für die Gesamtpopulation ist, wodurch zu positive oder zu negative Bewertungsergebnisse abgegeben werden. Dieses Problem kann durch eine größere Bewertungsgruppe verringert werden. Auch sollten besonders schlechte Bewertungen durch weitere Expertenkonsultationen überprüft werden. Hierbei ist die Erfahrung des Testverantwortlichen wichtig. Sollte ihm eine derartige Verfälschung der Ergebnisse auffallen, so sind entsprechende Maßnahmen zu ergreifen.

Ein weiteres Problem könnte ein Reihenfolgeeffekt (auch Primacy-Recency-Effekt genannt) [36] der zu bewertenden Gebrauchstauglichkeitsprobleme sein. Dieser Effekt besagt, dass die Reihenfolge der Bewertungen einen Einfluss auf deren Qualität haben kann. Bewertungen finden immer im Vergleich mit den vorherigen Bewertungen statt. So kann ein mittelmäßiges Problem schnell negativer bewertet werden, wenn vorher ein sehr geringes Problem bewertet wurde. Um diesen Effekt zu verringern, könnten vor einer Bewertung verschiedenschwere Beispielprobleme vorgestellt werden, wodurch die Evaluatoren für die Thematik sensibilisiert würden. Dies müsste jedoch in weiteren Studien untersucht werden.

In den durchgeführten Studien zeigt sich, dass häufig schwerwiegende und auch sicherheitskritische Probleme identifiziert werden. Eine Erklärung hierfür ist, dass viele Medizintechnikhersteller die Gebrauchstauglichkeit ihrer Geräte gar nicht, zu wenig oder zu spät betrachten. Ergebnisse aus Gebrauchstauglichkeitsevaluationen werden oft zu spät gewonnen, um sie noch ohne höheren Kostenaufwand in den laufenden Entwicklungsprozess integrieren zu können [202]. Gebrauchstauglichkeit und Nutzeranforderungen müssen bereits bei der Ideenfindung zu neuen Produkten berücksichtigt werden. Erst durch eine genaue Betrachtung und Evaluation der Gebrauchstauglichkeit und Einbeziehung der Benutzer in die Produktentwicklung ist ein gebrauchstaugliches Design möglich [226, 246, 305, 346]. Gebrauchstaugliche Geräte führen zu unmittelbaren Vorteilen für das medizinische Personal, das Krankenhaus und schließlich für die Patienten. Allerdings muss noch intensiv daran gearbeitet werden, um das Bewusstsein für die Bedeutung dieser Maßnahmen zu wecken [188].

Am Experimental-OP Tübingen wurde in Zusammenarbeit mit der Firma CardioBridge ein nutzerzentrierter Designprozess für das Bedienkonzept eines medizinischen Gerätes durchgeführt [44]. Hierfür wurde u.a. auch die in dieser Arbeit entwickelte Standardprozedur verwendet. Es zeigte sich, dass bestehende Designkonzepte auffallend verbessert werden konnten, und die Gebrauchstauglichkeit in Form von Effektivität, Effizienz und Zufriedenheit gesteigert wurde. Die neu entwickelte Bedienschnittstelle beurteilten die Nutzer als überaus positiv. Gerade von Fachexperten waren vermehrt Aussagen zu vernehmen, wie: „Wo ich das Gerät sehe, wird mir erst bewusst, wie kompliziert unsere Pumpe ist.“ (Fachkrankenschwester), „Das Gerät macht ein guten Eindruck und ist einfach und intuitiv.“ (Kardiotechniker), und: „Im Vergleich zu anderen Geräten ist die Menüführung positiv und sehr einfach.“ (Kardiologe).

Weiterhin ermöglicht die Betrachtung der Gebrauchstauglichkeit, besonders unter dem sicherheitskritischen Aspekt, wie in dieser Arbeit geschehen, eine neue Art der Fehleranalyse. Lange wurde für die Aufschlüsselung von Fehlerursachen ein individualistischer Ansatz verwendet, das heißt, dass beim Auftreten eines Fehlers sofort nach dem Fehler des einzelnen Bedieners gesucht wurde, selbst dann, wenn eine komplexe, unzuverlässige Technologie oder mehrere Personen eindeutig mitverantwortlich waren [118]. Es konnte gezeigt werden (vgl. Kap. 2.2., 5.1.4 u. 5.5), dass hierfür häufig Probleme in der Mensch-Maschine-Kommunikation verantwortlich sind. Die durch die Standardprozedur ermittelten Erkenntnisse können dazu verwendet werden, derartige Probleme und die daraus entstehenden Fehler zu verhindern oder zumindest zu verringern. Hierbei ist jedoch zu betonen, dass die Standardprozedur summativ die Gebrauchstauglichkeit bewertet und somit für serienreife Produkte oder späte Prototypen gedacht ist. Für die Gestaltung von Medizingeräten gibt es eine Vielzahl von Industrie- und Firmenstandards. Die vorliegende Arbeit setzt weder irgendeinen genormten Entwicklungsprozess (z.B. nach der DIN EN IEC 60601-1-6 oder DIN EN 62366) voraus noch behandelt sie alle verschiedenen Aktivitäten, die erforderlich sind, um eine gebrauchstaugliche Systemgestaltung sicherzustellen. Zwar können Teile dieser Arbeit als praktische Hilfsmittel für vorgeschriebene Prozesse verwendet werden, doch bildet sie in ihrer Gesamtheit eine Ergänzung zu vorhandenen Gestaltungsverfahren und liefert eine Möglichkeit, die Benutzerfreundlichkeit medizinischer Geräte im jeweiligen Kontext zu evaluieren und zu bewerten.

Von der Standardprozedur können verschiedene Interessensgruppen profitieren: Die Ergebnisse der Standardprozedur ermöglichen einem Hersteller medizinischer Geräte, Probleme in der Mensch-Maschine-Interaktion zu erkennen und zu beheben und gebrauchstaugliche Produkte zu entwickeln. Daneben kann er bei einem gebrauchstauglichen Produkt die Ergebnisse zu Marketingzwecken verwenden.

Anwendern und Betreibern von medizinischen Geräten ermöglichen die Ergebnisse der Standardprozedur, gebrauchstaugliche Geräte zu identifizieren und durch deren Auswahl die Effektivität, Effizienz, Sicherheit und Anwenderzufriedenheit zu steigern. Dies bedeutet eine nicht zu unterschätzende Hilfe bei der Anschaffungsentscheidung. Evaluatoren und Prüfinstituten bietet die Standardprozedur ein auf die internationalen Normen abgestimmtes Verfahren, welches den Evaluatoreffekt verringert und für die Gebrauchstauglichkeitsevaluation medizinischer Geräte eingesetzt werden kann. Benannte Stellen können die Ergebnisse der Standardprozedur dazu verwenden, eine Zertifizierung entsprechend der Normen DIN EN IEC 60601-1-6 und DIN EN IEC 62366 durchzuführen.

Ergonomische Erkenntnisse, die Mensch-Maschine-Kommunikation und das medizinische Arbeitssystem sind einem ständigen Wandel und Fortschritt ausgesetzt. Viele Bereiche sind durch Normen, Richtlinien und Gesetze geregelt, die demselben Wandel unterliegen. Aus diesem Grund soll der mit dieser Arbeit gegebene Bewertungsmaßstab für die Gebrauchstauglichkeit medizinischer Geräte lediglich ein Instrument sein, welches bei der Bewertung unterstützt. Für den Einzelfall hat der Anwender den Gebrauch der Methode mit den jeweiligen aktuellen Richtlinien und Normen abzugleichen und ihre Anwendbarkeit für die gegebene Fragestellung zu überprüfen.

Letztendlich bleibt festzuhalten, dass auch mit dem besten Design Menschen weiterhin Fehler machen werden. Die Norm IEC 1050 betont ausdrücklich, dass Ausbildung, Erfahrungen und Fähigkeiten das Risiko beeinflussen können, aber nicht dafür verwendet werden dürfen, um als Ersatz für eine Risikominderung zu dienen, welche durch eine konstruktive oder technische Schutzmaßnahme hätte erreicht werden können [94]. Zu dem geforderten Ziel, potentielle Fehlerquellen zu reduzieren, zählen auch die Aufdeckung von Gebrauchstauglichkeitsproblemen und die Zertifizierung gebrauchstauglicher Geräte in der Medizin. Erste Schritte hierzu sind bereits gemacht, jedoch wird es wohl noch ein langer Weg

sein, bis auch der Letzte verstanden hat, wie wichtig gebrauchstaugliche Produkte für die Patientensicherheit und für effektive Behandlungen sind.

*- Mistakes are a fact of life. It's the response to the error that counts -*  
N. Giovanni

## Zusammenfassung der Dissertation

Aktuellen Studien zufolge ist mangelnde Gebrauchstauglichkeit medizinischer Geräte eine häufige Ursache für Fehler und die daraus resultierenden Risiken und Gefahren für Personal und Patienten im medizinischen Arbeitsfeld. Gebrauchstauglichkeit ist somit ein zentrales Element für die Patientensicherheit und für effiziente Arbeitsabläufe. Durch eine neue Gesetzgebung und harmonisierte internationale Standards werden den Herstellern medizinischer Produkte Aktivitäten für die Entwicklung gebrauchstauglicher Medizinprodukte abverlangt. Explizite Methoden zur Erfassung und Bewertung der Gebrauchstauglichkeit medizinischer Produkte sind bisher jedoch noch nicht vorhanden. Traditionelle Forschungsansätze der Mensch-Maschine-Kommunikation sind essenziell, aber nicht ausreichend für die Komplexität typischer moderner Geräte im sicherheitskritischen Kontext der Medizin. Verfahren der Risikoabschätzung von Gebrauchstauglichkeitsproblemen in der Medizin existieren noch nicht.

Im Rahmen dieser Arbeit wird mit ausführlichen Studien eine Standardprozedur entwickelt, welche es ermöglicht, medizintechnische Produkte im Hinblick auf ihre Gebrauchstauglichkeit zu überprüfen und zu bewerten.

Basierend auf dem zweigeteilten Ansatz der Gebrauchstauglichkeitsevaluation und der Gebrauchstauglichkeitsbewertung kann die Gebrauchstauglichkeit medizinischer Produkte bestimmt werden. Hierfür verknüpft die Methode Elemente der Gebrauchstauglichkeitsevaluation mit Elementen des Risikomanagements. Erst durch diese Kombination kann die Gebrauchstauglichkeit unter dem Aspekt der Sicherheit betrachtet werden. Durch die standardisierte Vorgehensweise wird die Reproduzierbarkeit der Ergebnisse im Sinne von Reliabilität und Objektivität erreicht.

Für den Evaluationsteil der Standardprozedur wird nach theoretischen Herleitungen der Methodenelemente und entsprechenden Basisbausteinen in einem iterativen Prozess mit abwechselnden Entwicklungs- und Evaluationsphasen eine Methode entworfen. Durch vier Studien werden aufgestellte Thesen verifiziert und die Methode konkretisiert.

Für den Bewertungsteil der Standardprozedur werden als erstes die theoretischen Rahmenbedingungen für die Bewertung der Gebrauchstauglichkeit medizinischer Geräte hergeleitet. Hierzu werden die Bereiche der Mensch-Maschine-Interaktion und des Risikomanagements herangezogen, um in verschiedenen Bewertungsmöglichkeiten verknüpft zu werden. Analog zu der Entwicklung des Evaluationsteils werden wiederum in iterativen Entwicklungs- und Evaluationsphasen die geeigneten Methoden ausgewählt und konkretisiert. Das Ergebnis dieses Ansatzes ist die Bewertungsmethode UseProb.

In einer abschließenden Evaluationsphase werden die beiden Elemente der Standardprozedur zusammengeführt und evaluiert. Hierbei zeigt sich, dass die Standardprozedur die in dieser Arbeit gestellten Anforderungen und die Testgütekriterien Objektivität, Reliabilität und Validität erfüllt.



## Abbildungsverzeichnis

Abb. 1: Iterativer Entwicklungsleitfaden dieser Arbeit	7
Abb. 2: Allgemeines Wirkungsschema menschlicher Arbeit	9
Abb. 3: Interdisziplinarität der MMK in der Medizin	9
Abb. 4: Menschliche Informationsaufnahme und -ausgabe	10
Abb. 5: Phasen der Aktionstheorie nach Norman	10
Abb. 6: 3-Ebenen-Modell des menschlichen Verhaltens bei Bedientätigkeiten	13
Abb. 7: Einflüsse im medizinischen Arbeitssystem am Beispiel OP	14
Abb. 8: Das Seeheim-Modell	15
Abb. 9: Das IFIP Modell für Benutzerschnittstellen	16
Abb. 10: Das medizinische Arbeitssystem	19
Abb. 11: Verschiedene medizinische Geräte bis hin zum OP-System	25
Abb. 12: Benutzer medizinischer Geräte	29
Abb. 13: Abhängigkeiten im medizinischen Arbeitsumfeld	33
Abb. 14: Beispiel einer anthropometrischen Betrachtung	35
Abb. 15: Ein benutzerzentrierter Designprozess ...	38
Abb. 16: Usability Engineering Lifecycle nach Mayhew	39
Abb. 17: Ein Hochfrequenz-Chirurgiegerät benötigt zusätzliche Informationen	45
Abb. 18: Kosteneinsparungen durch Usability Engineering	48
Abb. 19: Kosten im Entwicklungsprozess	50
Abb. 20: Fehlertypen unterschiedlicher Medizinprodukte	54
Abb. 21: Steckverbindung mit Risikopotential	55
Abb. 22: Entwicklungsprozess für gebrauchstaugliche medizinische Produkte	57
Abb. 23: Vorteile eines benutzerzentrierten Entwicklungsprozesses	58
Abb. 24: Operationalisierung der Gebrauchstauglichkeit...	75
Abb. 25: Zusammenhang zwischen der DIN EN IEC 60601-1-6 und ...	76
Abb. 26: Unterschiedliche Ausprägungen eines Prototypen...	80
Abb. 27: Gefundene Probleme und Kosten-Nutzen-Verhältnis...	82
Abb. 28: Typischer Nutzertest im Experimental-OP in Tübingen	89
Abb. 29: Prozentzahlen der gefundenen Gebrauchstauglichkeitsprobleme...	89
Abb. 30: Anteil der gefundenen Gebrauchstauglichkeitsprobleme ...	101
Abb. 31: Beziehung zwischen dem bewerteten Nutzen...	102
Abb. 32: Gefundene Probleme durch eine Heuristische Evaluation...	105
Abb. 33: Entwicklungsprozess der Standardprozedur...	108
Abb. 34: Entscheidungsbaum für Effizienzminderung	114
Abb. 35: Zentrale Bestandteile des medizinischen Nutzungskontextes	128
Abb. 36: Beispielausschnitt einer HTA für ein Hochfrequenzchirurgiegerät	133
Abb. 37: Vom Nutzungskontext zu den Evaluationskriterien	137
Abb. 38: Beispielfragen zur Nutzerzufriedenheit, adaptiert nach dem PSSUQ	141
Abb. 39: Einstufung der Zufriedenheit zu einzelnen Nutzeraufgaben	141
Abb. 40: Vergleichendes Ranking zur Ermittlung der Nutzerzufriedenheit	142
Abb. 41: Szenario und Aufgaben einer Gebrauchstauglichkeitsprüfung	152
Abb. 42: Beispiel eines Aufgabenszenarios	153
Abb. 43: Der Experimental-OP Tübingen...	155
Abb. 44: Abschnitte eines standardisierten Berichts...	156
Abb. 45: Regressionsgerade und Bland-Altman-Plott...	158
Abb. 46: Hochfrequenzchirurgiegeräte der Gebrauchstauglichkeitsstudie ...	163

Abb. 47: Beispiel aus dem Vergleichsfragebogen...	165
Abb. 48: Verschiedene Antworten auf die Frage, wie oft ...	166
Abb. 49: Berechnung des angepassten p-Werts (Adj p) für das HF-Gerät 1	169
Abb. 50: Berechnung der Fehlerfindungsrate...	169
Abb. 51: Hierarchische Aufgabenanalyse für den ErgoGrip®	174
Abb. 52: Beispiele für die Zufriedenheit mit Aktionen von...	177
Abb. 53: Fragebogen und Ergebnisse des vergleichenden Fragebogens...	178
Abb. 54: Berechnung der Fehlerfindungsrate...	178
Abb. 55: Beispiel für die Auswertung des ASQs...	180
Abb. 56: Untersuchungsgegenstand der Spritzenpumpenstudie	182
Abb. 57: Testumgebung: Die Ziffern 1-4 zeigen die Kamerapositionen ...	184
Abb. 58: Aufgabenbearbeitungsdauer mit Standardabweichung...	185
Abb. 59: Gebrauchstauglichkeitsprobleme aus Experteninspektion (EI) ...	186
Abb. 60: Berechnung der Fehlerfindungsrate...	187
Abb. 61: Schematische Darstellung der in diesem Kapitel hergeleiteten...	189
Abb. 62: Entwicklungsprozess der Standardprozedur...	202
Abb. 63: Schweizer-Käse-Modell nach Reason, bezogen auf...	203
Abb. 64: Menschliche Verarbeitungsebenen mit typischen Fehlern	207
Abb. 65: Klassifikation von Handlungen adaptiert nach der ...	212
Abb. 66: UAF-Interaktionszyklus und Zusammenhang zwischen ...	213
Abb. 67: Ursachen, Mechanismen und Auswirkungen menschlicher Fehler...	215
Abb. 68: Fehler im Interaktionszyklus	215
Abb. 69: Systemhierarchie menschlicher Fehler in der Medizin	217
Abb. 70: Zusammenwirken der Klassifizierungen medizinischer Fehler nach ...	218
Abb. 71: Das Modell nach Rasmussen, adaptiert für den OP-Arbeitsplatz	221
Abb. 72: Klassifizierung gerätebezogener Fehler	223
Abb. 73: Fehlerklassifizierungsschema nach der Ursache...	224
Abb. 74: Zusammenhang zwischen Sicherheit und Risiko	225
Abb. 75: Zusammenhang sicherheitsbezogener Begriffe...	226
Abb. 76: Risikoelemente	230
Abb. 77: Iteratives Verfahren von Risikobeurteilung und Risikominderung	232
Abb. 78: Beispiel für einen einfachen Fault Tree	236
Abb. 79: Beispiel einer Risikomatrix	243
Abb. 80: Allgemeines Prinzip eines Risikographs	245
Abb. 81: Beispiel eines Risikodiagramms mit drei Bereichen ...	246
Abb. 82: Abhängigkeit der Geräteanwendung vom Nutzungskontext...	250
Abb. 83: Methoden des Usability Engineerings im Risikomanagement	250
Abb. 84: Kombination der Gebrauchstauglichkeitsbewertung und ...	264
Abb. 85: 3-dimensionale Matrix zur Bewertung der Gebrauchstauglichkeit...	267
Abb. 86: Risikograph zur Bewertung der Gebrauchstauglichkeit...	273
Abb. 87: Szenarienbeschreibung und Bewertungsbereich	277
Abb. 88: Zusammenfassung der Bewertung der evaluierten Szenarien ...	279
Abb. 89: Expertenzufriedenheit mit der entwickelten Methode ...	280
Abb. 90: Bewertung des 1. HF-Gerätes in drei Sitzungen...	288
Abb. 91: Bewertung des 2. HF-Gerätes in drei Sitzungen...	288
Abb. 92: Bewertung des 3. HF-Gerätes in drei Sitzungen...	289
Abb. 93: Problembewertung durch fünf Experten	293
Abb. 94: Ermittelte Gebrauchstauglichkeitsprobleme aus zwei ...	301
Abb. 95: Problembewertung zweier Studien im Vergleich	302

Abb. 96: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug ...	303
Abb. 97: Bland-Altman-Methode zur Visualisierung des Zusammenhangs...	303
Abb. 98: Ergebnisse der Problembewertung zweier wiederholender Studien	304
Abb. 99: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug ...	305
Abb. 100: Bland-Altman-Methode zur Visualisierung des ...	305
Abb. 101: Bewertungsergebnisse der zweiten Studie (y-Achse) in Bezug ...	317
Abb. 102: Bland-Altman-Methode zur Visualisierung des Zusammenhangs ...	317

## **Tabellenverzeichnis**

Tab. 1: Medizinische Primär- und Sekundäraufgaben	23
Tab. 2: Risikoklassen von Medizinprodukten	27
Tab. 3: Exemplarischer Überblick über Gebrauchstauglichkeitsdefinitionen	34
Tab. 4: Aussagen zur Gerätebedienung von Chirurgen und OP-Pflegekräften	43
Tab. 5: Fehlerwahrscheinlichkeit in Abhängigkeit von Stress und ...	44
Tab. 6: Nutzen durch gebrauchstaugliche Geräte für Anwender und Hersteller	52
Tab. 7: Gebrauchstauglichkeitskriterien mit Definitionen	74
Tab. 8: Verschiedene Methoden des Usability Engineerings...	78
Tab. 9: Vergleich von drei Evaluationsmethoden nach Desurvire et al.	104
Tab. 10: Relevanz bekannter Prüfmethode für die ...	122
Tab. 11: Beispiel für eine Nutzer/Aufgaben Matrix	130
Tab. 12: Zusammenfassung einiger Komponenten des Nutzungskontextes...	136
Tab. 13: Beispiele für Maße der Gebrauchstauglichkeit	138
Tab. 14: Anzahl der Testnutzer in Abhängigkeit von...	149
Tab. 15: Gefundene Probleme aus Heuristischer Evaluation (HE) und ...	167
Tab. 16: Identifizierte Probleme in einem Einzeltest und ...	170
Tab. 17: Gefundene Probleme bei einem Handgriff durch ...	177
Tab. 18: Beispiel für die Zeitmessung einer Aufgabe	179
Tab. 19: Beispiele für die Fehlermessung	179
Tab. 20: Bei den durchgeführten Studien durch...	185
Tab. 21: Klassifizierung der Slips (Fehler) nach vermuteten Ursachen	207
Tab. 22: Fehlerklassifizierungsschema nach Rasmussen	209
Tab. 23: Beispiele von Gebrauchstauglichkeitsproblemen ...	211
Tab. 24: Fehlergegenmaßnahmen in den einzelnen Phasen der Interaktion	216
Tab. 25: Eine handlungsorientierte Taxonomie für Versehen...	220
Tab. 26: PHA am Beispiel einer Spritzenpumpe	237
Tab. 27: Exemplarische PAAG am Beispiel einer Spritzenpumpe	238
Tab. 28: Einstufung des Schweregrades der Fehlerkonsequenzen	241
Tab. 29: Wahrscheinlichkeitseinteilung nach der DIN EN 61508-5	242
Tab. 30: Beispiel einer Risikomatrix aus der DIN EN 61508	243
Tab. 31: Traditionelles Risikomanagement vs. Human Factors Analysis	251
Tab. 32: Beispiele für die Fehlersuche	255
Tab. 33: Ranking der Auftretenswahrscheinlichkeit eines ...	257
Tab. 34: Vier-Ebenen-Ranking zur Bewertung des Schweregrades von ...	257
Tab. 35: Fünf-Ebenen-Ranking zur Bewertung des Schweregrades von ...	258
Tab. 36: Vier-Ebenen-Ranking zur Bewertung des Schweregrades von ...	258
Tab. 37: Drei-Ebenen-Ranking zur Bewertung des Schweregrades von ...	258

Tab. 38: Zertifizierungsstufen für die Folgen von Ausfällen im Flugverkehr	259
Tab. 39: Bewertungsstufen nach Schmidtke	260
Tab. 40: Fehlerkategorien nach dem entstandenen Schade	261
Tab. 41: Einstufung der Auftretenswahrscheinlichkeit eines ...	264
Tab. 42: Einstufung einer Arbeitsaufgabe und eines ...	265
Tab. 43: 4-stufige Kategorisierung nach dem potentiellen Schaden	266
Tab. 44: Bewertungskonzept der 3-D Matrix (n/a = nicht anwendbar).	267
Tab. 45: Bewertungsstufen für die Gebrauchstauglichkeit medizinischer ...	268
Tab. 46: Faktoren für die Klassifizierung von ...	269
Tab. 47: Gewichtungsfaktoren	270
Tab. 48: Ergebnisstufen für das 2. Konzept	270
Tab. 49: Risikoparameter des Risikographens	273
Tab. 50: Teilnehmer der 2. Entwicklungsstudie	278
Tab. 51: Bewertungen der Gebrauchstauglichkeitsprobleme ...	285
Tab. 52: Bewertung der Gebrauchstauglichkeit durch Diskussion	286
Tab. 53: Bewertung der Gebrauchstauglichkeit durch Diskussion und ...	287
Tab. 54: Gewählte Testaufgaben der identifizierten Aufgaben; ...	300

## Abkürzungsverzeichnis

Abb.	Abbildung
ALARP	As low as reasonably practicable
ATA	Anästhesietechnischer Assistent
BfArM	Bundesinstitut für Arzneimittel und Medizinprodukte
CWT	Cognitive Walkthrough
DeGEval	Deutsche Gesellschaft für Evaluation
DGCH	Deutsche Gesellschaft für Chirurgie
DIN	Deutsche Institut für Normung
DRG	Diagnosis Related Groups
EMG	Elektromyografie
EXOP	Experimental OP
FDA	Food and Drug Administration
FMEA	Failure Mode and Effects Analysis
FTA	Fault Tree Analysis
HAZOP	Hazard and Operability
HE	Heuristische Evaluation
HF	Hochfrequenz (Hohfrequenzchirurgie)
HTA	Hierarchical Task Analysis
IEC	International Electrotechnical Commission (Normungsgremium für Elektrotechnik)
IFIP	International Federation for Information Processing
IOM	Institute of Medicine
ISO	International Organization for Standardization (Internationale Organisation für Normung)

KANN	Kommission für Arbeitsschutz und Normung
MDD	Medical Device Directive (Richtlinie 93/42/EWG über Medizinprodukte)
MMI	Mensch-Maschine-Interaktion
MMK	Mensch-Maschine-Kommunikation
MPG	Medizinproduktegesetz
MTBF	Mean Time Between Failures
NT	Nutzertest
NWA	Nutzwertanalyse
OHA	Operating Hazard Analysis
OP	Operationssaal
OTA	Operationstechnischer Assistent
PAAG	Verfahren für Prognose, Auffinden der Ursache, Abschätzen der Auswirkungen, Gegenmaßnahmen
PHA	Preliminary Hazard Analysis
Tab.	Tabelle
UT	Usability Test
VDE	Verband der Elektrotechnik, Elektronik und Informationstechnik e.V.

## Literaturverzeichnis

- 1 Algedri J., Frieling E. (2001): Human FMEA. Fachbuchverlag Leipzig
- 2 Alvarado C.J. (2007): The physical environment in health care. In: Carayon P.: Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Lawrence Erlbaum Associates, Hillsdale. NJ S.287-307
- 3 Andre T.S., Hartson H.R., Belz S.M., McCreary F.A. (2001): The user action framework: a reliable foundation for usability engineering support tools. In: International Journal of Human-Computer Studies, 54, (2001) S. 107-136
- 4 Annett J.(2003): Hierarchical Task Analysis. In: Diaper D., Stanton N. A.: The Handbook of Task Analysis for Human-Computer Interaction. Edited by Bournemouth University. Lawrence Erlbaum Associates. S.67-82
- 5 ANSI-NCITS 354-2001(2001): Common industry format for usability test reports, Washington DC. American National Standards Institute
- 6 Ansorge P., Frick G., Friedrich J., Haupt U. (1999): "Ergonomie geprüft" – Das Ende der Benutzungsprobleme? In:IM - Die Fachzeitschrift für Information Management & Consulting, 3/1999 (14. Jahrg., August 1999), S. 59-64
- 7 Antoni C. (1994): Gruppenarbeit mehr als ein Konzept In: Antoni C. (Hrsg.): Gruppenarbeit im Unternehmen. Konzepte, Erfahrungen, Perspektiven, Beltz, Weinheim S.19-48
- 8 Arnold B., Roe R. A. (1987): User errors in human-computer interaction. In: Frese M., Ulich E., Dzida W. (Hrsg.): Psychological issues of human computer interaction in the workplace. Amsterdam. Elsevier. S.203-220
- 9 Arzt & Wirtschaft, 10/2007, S. 13
- 10 Aspden P., Corrigan J.M., Wolcott J., Erickson S.M. (2003): Patient Safety: Achieving a New Standard of Care. Washington, D.C., Institute of Medicine of the National Academies. National Academies Press
- 11 Backhaus C. (2004): Entwicklung einer Methode zur Analyse und Bewertung der Gebrauchstauglichkeit von Medizintechnik. PROMEDIKS - Prozessorientierte Medizintechnik in klinischen Systemen. Dissertation der Technischen Universität Berlin
- 12 Bailey B. (2001): Heuristic Evaluations vs. Usability Testing. Abruf unter: [www.webusability.com/article\\_heuristic\\_evaluation\\_part1\\_2\\_2001.htm](http://www.webusability.com/article_heuristic_evaluation_part1_2_2001.htm) (14.05.2009)
- 13 Bailey R. W., Allan R. W., Raiello P. (1992): Usability testing vs. heuristic evaluation: A head-to-head comparison. In: Proceedings of the Human Factors Society 36th annual meeting, 409-412
- 14 Baker G.R., Norton P.G., Flintoft V., Blais R., Brown A., Cox J., Etchells E., Ghali W.A., Hébert P., Majumdar S.R., O'Beirne M., Palacios-Derflingher L., Reid R.J., Sheps S., Tamblyn R. (2004): The Canadian Adverse Events Study: the incidence of adverse events among hospit. In: 2004 May 25;170(11):1678-86
- 15 Basnyat S., Navarre S., Palanque P. (2005): Complexity of Design in Safety Critical Interactive Systems: Gathering, Refining, Formalizing Multi-Type and Multi-Source Information while Ensuring Consistency, Reliability, Efficiency and Error-Tolerance. 2nd Workshop on Complexity in Design and Engineering. Glasgow, Scotland. UK 10th-12th March 2005. S. 45-58
- 16 Bauer H.(2006): Arbeitsplatz OP: Realität und Anspruch. In: Deutsches Ärzteblatt, Jg. 103 Heft 47, 24.
- 17 Baumann T. (2007): Usability-Studie zur Untersuchung der Gebrauchstauglichkeit von Hochfrequenz-Chirurgie Geräten. Diplomarbeit im Fachbereich Medizintechnik der Hochschule Ulm

- 18 Bayerisches Staatsministerium für Wirtschaft, Infrastruktur, Verkehr und Technologie (2005): Merkblatt zur EU-Richtlinie 93/42/EWG
- 19 Beaudouin-Lafon M., Mackay W. (2008): Prototyping Tools and Techniques. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application. New York. Taylor & Francis. S. 1017-1039
- 20 Becks T. (2005 ): Trends in der Medizintechnik. In: Forum Forschung. Universität Duisburg. Abruf unter: <http://www.forum-forschung.de/2004/pdf/fofo20042005.pdf> (14.05.2009)
- 21 Bernotat R.(1966): Die Anthropotechnik als wissenschaftliche Disziplin. DGRR/WGLR Jahrestagung: Bad Godesberg
- 22 Bevan N. (1995): Measuring Usability as Quality of Use. In: Journal of Software Quality Issue, Vol.4/1995. S.115-140
- 23 Bevan N., Macleod M. (1994):Usability measurement in context. In: Behaviour and Information Technology, 13, National Physical Laboratory, Teddington, Middlesex, UK. S.132-145
- 24 Beywl W. (2002): Standards für Evaluation. Köln. Deutsche Gesellschaft für Evaluation e.V.
- 25 BfArM (2009): Statistische Auswertung der im Zeitraum 01.01.2005 bis 31.12.2008 abschließend bewerteten Risikomeldungen. Eingetretene Auswirkungen in Verbindung mit einer Meldung. Abruf unter: [http://www.bfarm.de/cln\\_042/nn\\_424526/DE/Medizinprodukte/riskinfo/wissauf/statist-auswertung.html](http://www.bfarm.de/cln_042/nn_424526/DE/Medizinprodukte/riskinfo/wissauf/statist-auswertung.html) (10.03.2008)
- 26 Bias R. G.(1994): The Pluralistic Usability Walkthrough: Coordinated Emphathies. In: Nielsen J., Mack R. (Eds.): Usability Inspection Methods, John Wiley and Sons, New York, S. 63-76
- 27 Bildschirmarbeitsverordnung (BildscharbV): Verordnung über Sicherheit und Gesundheitsschutz bei der Arbeit an Bildschirmgeräten.
- 28 BKK Bundesverband (2005):BKK Faktenspiegel. Abruf unter: [http://www.bkk-achenbach.de/news/bkk\\_faktenspiegel/2005/FS\\_09\\_05.pdf](http://www.bkk-achenbach.de/news/bkk_faktenspiegel/2005/FS_09_05.pdf) (01.11.2008)
- 29 Bland J.M., Altman D.G. (1986): Statistical methods for assessing agreement between two methods of clinical measurement. In: Lancet. 1986 Feb 8;1. S.307-310.
- 30 Bleyer S. (1992): Medizinisch-technische Zwischenfälle in Krankenhäusern und ihre Verhinderung. In: Anna O., Hartung C. (Hrsg.): Mitteilungen des Instituts für Biomedizinische Technik und Krankenhaustechnik der Medizinischen Hochschule Hannover. Hannover. Fachverlag für Krankenhaustechnik
- 31 Böckmann R.D., Frankenberger H. (1994): Durchführungshilfen zum Medizinproduktegesetz. Köln. TÜV-Reinland
- 32 Boggian K., Schlegel M. (2008): Risiko und Nebenwirkungen für Patient und Gesellschaft. Es gibt keine Packungsbeilage zum Thema Arzt. Vortrag auf dem 13. St. Galler Infekttag, St. Gallen
- 33 Bogner M.S. (1998): An Indroduction to design, evaluation, and usability testing. In Rice V. J. (Eds.): Ergonomics in Health Care and Rehabilitation. Boston. Butterworth-Heinemann S. 231-47
- 34 Booth P. (1991): Errors and theory in human-computer interaction. In: Acta Psychologica (78), S.69-96.
- 35 Börcsök J. (2006): Funktionale Sicherheit- Grundzüge sicherheitstechnischer Systeme. Hüthig GmbH & Co. KG, Heidelberg
- 36 Bortz J., Döring N.(2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg. Springer Medizin Verl.
- 37 Bös K. (Hrsg.) (2001). Handbuch motorische Tests. Göttingen, Hogrefe

- 38 Brackemann H.(2007): Die Untersuchung der Benutzungsfreundlichkeit in vergleichenden Warentests. Vortrag auf dem 8. World Usability Day. Stuttgart
- 39 Brannen V., Henning M., Jacobsen J., Korff J.J., Lonczewski F., Pöthig T., Zimmermann G.: Was Sie schon immer über Usability erfahren wollten. Stuttgart. German Chapter of the Usability Professionals
- 40 Brooke J.(1996): A "quick and dirty" usability scale. In: Jordan P. W. , Thomas B. , Weerdmeester B. A. , McClelland A. L. (Eds.): Usability Evaluation in Industry. London. Taylor and Francis
- 41 Büchel D. (2005): Entwicklung und Visualisierung weitergehender Szenarien für ein Fahrerinformationssystem im Pkw. Diplomarbeit, Stuttgart. DaimlerChrysler AG u. Technische Universität Ilmenau. Fakultät für Elektrotechnik und Informationstechnik, Institut für Medientechnik
- 42 Büchel D., Baumann T., Matern U. (2007): Usability of Radofrequency Devices in Surgery. In: Holzinger A (Eds.): HCI and Usability for Medicine and Health Care. USAB 2007, LNCS. S 97-104
- 43 Büchel D., Hallabrien B., Marvik, Matern U.: Ergonomics of disposable handles for minimally invasive surgery. Accepted for publication in Surgical Endoscopy.
- 44 Büchel D., Scherrer M., Matern U. (2009): Benutzerzentrierte Entwicklungsprozesse für ein nicht täglich verwendetes Notfallgerät. In: Hermeneit A., Steffen A., Stockhardt J. (Hrsg.): Medizinprodukte planen, entwickeln, realisieren (Begleit-CD), Köln, TÜV Media GmbH
- 45 Büchel D., Schiendzielorz E., Decker K, Matern U.: Sicherheitskritische Aspekte der Gebrauchstauglichkeit von Spritzenpumpen. Bisher unveröffentlicht.
- 46 Büchel D., von Spreckelsen H. (2005): Usability Testing von interaktiven Fernseh Anwendungen in England. Brighton, Ilmenau. Medienprojekt an der Technische Universität Ilmenau. Fakultät für Elektrotechnik und Informationstechnik, Institut für Medientechnik
- 47 Bullinger H.-J.(1994): Ergonomie, Produkt- und Arbeitsplatzgestaltung. Stuttgart. Teubner Verlag
- 48 Bullinger H.-J., Muntzinger W.F., Schindhelm R.(1990): Methodische Ansätze der ergonomischen Arbeitsmittelgestaltung. In: Friesdorf W., Schwilk B., Hähnel J. (Hrsg.): Ergonomie in der Intensivmedizin. Biblomed. S.79-94
- 49 Bundesärztekammer (2006): Stellungnahme der Zentralen Kommission zur Wahrung ethischer Grundsätze in der Medizin und ihren Grenzgebieten bei der Bundesärztekammer zur Ethikberatung in der klinischen Medizin: Deu. Ärzteblatt; Jg. 103; Heft 24; S. 1703 7
- 50 Çakir A. (2000): Usability Engineering - Vom Forschungsobjekt zur Technologie. In: HMD - Praxis Wirtschaftsinform 212: 7-21
- 51 Canton J.(2006): Vortrag auf dem Trendtag 2006 in Hamburg. Zitiert in: Krömker H. (2006):Usability - Stand der Forschung. In: Hennig J., Tjarks-Sobhani M. (Hrsg.):Usability und Technische Dokumentation. Lübeck. Schmidt-Römhild
- 52 Card S.K., Moran T.P., Newell A. (1983): The Psychology of Human-Computer Interaction. New Jersey, Lawrence Erlbaum.
- 53 Carroll J. M. (2000): Making use: Scenario-based design for human-computer interactions. Cambridge, MA: MIT Press.
- 54 Carroll J. M., Mack R. L. (1984): Learning to use a word processor: By doing, by thinking, and by knowing. In: Thomas J. C., Schneider M. L. (Eds.): Human factors in computer systems. Norwood, NJ: Ablex Publishing. S.13-51
- 55 Carroll J. M., Rosson M. B.(1990): Human-computer interaction scenarios as a design representation.In: Proceedings of the 23rd Annual Hawaii International Conference on Systems Sciences. Los Alamitos, CA: IEEE Computer Society Press S. 555-561



- 56 Carroll R., Carroll J.M.(2002): Scenario-based design In: Jacko J.A., Sears A. (Hrsg.): The human-computer interaction Handbook – Fundamentals, Evolving, Technologies and Emerging Applications. Mahwah. NY. Lawrence Erlbaum Association. Inc. Publishers.
- 57 Cebulla M. (2003): Kooperation und Sicherheit in soziotechnischen Systemen, Technische Universität Berlin. Abruf unter: [http://uebb.cs.tu-berlin.de/kosis/pr\\_kosis.html](http://uebb.cs.tu-berlin.de/kosis/pr_kosis.html) (28.10.2008)
- 58 Chalmers P. A. (2003): The role of cognitive theory in human-computer interface. In: Computers in Human Behavior, 19 (5), S.593-607
- 59 Chang A., Schyve P. M., Croeau R.J., O’Leary D. S., Jerod M. L (2005): The JCAHO patient safety event taxonomy: a standardized terminology and classification schema for near misses and adverse events. In: Int J Qual Health Care 2005;17(2):95-105.
- 60 Chantler C. (1999): The role and education of doctors in the delivery of health care. Lancet (1999) 353. S. 1178–81.
- 61 Chin J. P., Diehl V. A., Norman K. L. (1988): Development of an instrument measuring user satisfaction of the humancomputer interface. In: Proceedings of SIGCHI ‘88 New York: ACM/SIGCHI S. 213– 218
- 62 Chisnell D., Redish J. (2005): Designing Web Sites for Older Adults: Expert Review of Usability for Older Adults at 50 Web Sites. Association for the Advancement of Retired Persons (AARP). Abruf unter: [aarp.org/olderwiserwired](http://aarp.org/olderwiserwired) (03.04.2008)
- 63 Cina J. L., Gandhi T. K., Churchill W., Fanikos J., McCrea M., Mitton P., Rothschild J. M., Featherstone E., Keohane C., Bates D. W., Poon E. G. (2006): How Many Hospital Pharmacy Medication Dispensing Errors Go Undetected? In: Journal on Quality and Patient Safety Vol. 31 No. 2 S.73 – 80
- 64 Clark R. E., Feldon D., vanMerrienboer J., Yates K., Early S.(2008): Cognitive Task Analysis. In: Spector J.M., Merrill M.D., van Merrienboer J.J.G., Driscoll M.P. (Eds.). Handbook of research on educational communications and technology (3rd ed.). Mahwah NJ, Lawrence Erlbaum Associates.
- 65 Cockton G., Woolrych A., Lavery D. (2008): Inspection-Based Evaluations. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application, New York. Taylor & Francis S. 1171-1190
- 66 Cordes J., Dornheim J., Preim B. (2009): Szenariobasierte Entwicklung von Systemen für Training und Planung in der Chirurgie. In: ICOM - Zeitschrift für interaktive und kooperative Medien 1/2009 S.5-9
- 67 Courage C., Redish J., Wixon D. (2008): Task Analysis. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application, New York. Taylor & Francis. S. 937-947
- 68 Cox T., Randall R., Griffiths A. (2002): Interventions to control stress at work in hospital staff. Norwich. The Institute of Work, Health and Organisations for the Health and Safety Executive.
- 69 Crystal A, Ellington B. (2004): Task analysis and human-computer interaction: approaches, techniques, and levels of analysis. In: Proceedings of the Tenth Americas Conference on Information Systems, New York,
- 70 Dahm M. (2006): „Grundlagen der Mensch-Computer-Interaktion. New York. Pearson Studium
- 71 DEF 00-56-2 (1995): Safety management requirements for defence systems containing programmable electronics, Great Britain. Ministry of Defence. Directorate of Standardization., MoD (Glasgow)
- 72 Desurvire H. (1994): Faster, Cheaper: Are Usability Inspection Methods as Effective as Empirical Testing? In: Nielsen J., Mack R.L.: (Eds.): Usability Inspection Methods. New York: John Wiley & Sons, S.173-202

- 73 Desurvire H. W., Kondziela J. M., Atwood M. E. (1992): What is gained and lost when using evaluation methods other than empirical testing. In: Monk A., Diaper D., Harrison M. D. (Eds.): People and computers VII. Cambridge, UK: Cambridge University Press. S. 89-102.
- 74 Desurvire H., Lawrence D., Atwood M.(1991): Empiricism versus judgment: Comparing user interface evaluation methods on a new telephone-based interface. In: ACM SIGCHI Bulletin, 23(4), S.58-59
- 75 Deutsche Akkreditierungsstelle Technik GmbH (2004). DATech–Prüfhandbuch: Usability–Engineering–Prozess (Version 1.3). Frankfurt a.M.: Deutsche Akkreditierungsstelle Technik e.V.
- 76 Deutsche Akkreditierungsstelle Technik GmbH (2007): Leitfaden Usability. Frankfurt/Main. Deutsche Akkreditierungsstelle Technik e.V.
- 77 Deutsche Akkreditierungsstelle Technik GmbH (2008): DATech-Prüfhandbuch Gebrauchstauglichkeit. Leitfaden für die ergonomische Evaluierung von interaktiven Systemen auf Grundlage von DIN EN ISO 9241, Teile 11 und 110. Version 3.4. Abruf unter: [www.datech.de](http://www.datech.de) (12.03.2008)
- 78 Deutsche Gesellschaft für Biomedizinische Technik im VDE (DGBMT) (2003): Ergonomie in der Medizintechnik - Potenzial zur Qualitätsverbesserung in der medizinischen Versorgung Frankfurt am Main
- 79 Deutsche Gesellschaft für Qualität e.V. (2008): FMEA - Fehlermöglichkeits- und Einflussanalyse. 4. Auflage. Frankfurt, Berlin, Wien, Zürich: Beuth Verlag GmbH
- 80 DGBMT-Fachausschuss Ergonomie in der Medizintechnik (2004): Ergonomie in der Medizintechnik: Potenzial zur Qualitätsverbesserung in der medizinischen Versorgung. Deutsche Gesellschaft für Biomedizinische Technik im VDE (DGBMT). Frankfurt am Main. (2004)
- 81 Dhillon B.S.(2000) Medical Device Reliability and Associated Areas. Boca Ration FL. CRC Press
- 82 Diaper D. (1989): Task observation for Human-Computer Interaction. In: Diaper D. (Eds.), Task analysis for human computer interaction Chichester, England: Ellis Horwood. S. 210–237
- 83 Diaper D., (2004): Understanding Task Analysis for Human-Computer Interaction. In: Diaper D., Stanton N. A.: The Handbook of Task Analysis for Human-Computer Interaction. Mahwah, New Jersey. Edited by Bournemouth University. Lawrence Erlbaum Associates. S. 5-48
- 84 Diaper D., Stanton N. A.(2003): Understanding Task Analysis for Human-Computer Interaction. Stanton Bournemouth University In: Diaper D., Stanton N. A.: The Handbook of Task Analysis for Human-Computer Interaction. Edited by Bournemouth University. Lawrence Erlbaum Associates. S. 603-619
- 85 DIN 25424 (1990): Fehlerbaumanalyse; Handrechenverfahren zur Auswertung eines Fehlerbaumes. Berlin. Beuth
- 86 DIN 31000 (1979): Allgemeine Leitsätze für das sicherheitsgerechte Gestalten technischer Erzeugnisse. Berlin. Beuth
- 87 DIN 33402 (2008) Körpermaße des Menschen. Berlin. Beuth
- 88 DIN 33408 (2008): Körperumrißschablonen. Berlin. Beuth
- 89 DIN 33411 (1982) Körperkräfte des Menschen. Berlin. Beuth
- 90 DIN 33419 (1993): Ergonomische Prüfung von Produktentwürfen und Industrieerzeugnissen. Berlin. Beuth
- 91 DIN 33430 (2002): Anforderungen an Verfahren und deren Einsatz bei berufsbezogenen Eignungsbeurteilungen. Berlin. Beuth

- 
- 92 DIN 66050 (1980): Gebrauchstauglichkeit – Begriff. Berlin. Beuth
  - 93 DIN 66054 (1982): Warentest; Grundsätze für die technische Durchführung. Berlin. Beuth
  - 94 DIN EN 1050 (1997): Sicherheit von Maschinen - Leitsätze zur Risikobeurteilung. Berlin. Beuth
  - 95 DIN EN 13861 (2003): Sicherheit von Maschinen - Leitfaden für die Anwendung von Ergonomie-Normen bei der Gestaltung von Maschinen. Berlin. Beuth
  - 96 DIN EN 60601-1-4 (2003): Medizinische elektrische Geräte: Allgemeine Festlegungen für die Sicherheit - Ergänzungsnorm: Programmierbare elektrische medizinische Systeme. Berlin. Beuth
  - 97 DIN EN 60601-1-6 (2005): Medizinische elektrische Geräte Teil 1-6: Allgemeine Festlegungen für die Sicherheit - Ergänzungsnorm: Gebrauchstauglichkeit. Berlin. Beuth
  - 98 DIN EN 61025 (2007): Fehlzustandsbaumanalyse. Berlin. Beuth
  - 99 DIN EN 614-1 (2009): Ergonomische Gestaltungsgrundsätze – Teil 1 Begriffe und allgemeine Leitsätze. Berlin. Beuth
  - 100 DIN EN 61508 Bbl 1 (VDE 0803 Bbl 1): 2005-10 - Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbar elektronischer Systeme. Beiblatt 1. Berlin. Beuth
  - 101 DIN EN 61508-4 (2006): Funktionale Sicherheit elektrischer/elektronischer/programmierbar elektronischer sicherheitsbezogener Systeme - Teil 4: Begriffe und Abkürzungen. Berlin. Beuth
  - 102 DIN EN 61508-5 (2008): Funktionale Sicherheit sicherheitsbezogener elektrischer/elektronischer/programmierbarer elektronischer Systeme - Teil 5: Beispiele von Methoden für die Bestimmung von Sicherheits-Integritätsleveln. Berlin. Beuth
  - 103 DIN EN 61511 (2005): Funktionale Sicherheit - Sicherheitstechnische Systeme für die Prozessindustrie. Berlin. Beuth
  - 104 DIN EN 62366 (2008): Medizinprodukte - Anwendung der Gebrauchstauglichkeit auf Medizinprodukte. Berlin. Beuth
  - 105 DIN EN ISO 13407 (1999): Benutzer-orientierte Gestaltung interaktiver Systeme. Berlin. Beuth
  - 106 DIN EN ISO 13485 (2003): Medizinprodukte – Qualitätsmanagementsysteme – Anforderungen für regulatorische Zwecke. Berlin. Beuth
  - 107 DIN EN ISO 13849-1 (2009): Sicherheit von Maschinen. Sicherheitsbezogene Teile von Steuerungen. Teil 1. Allgemeine Gestaltungsleitsätze. Berlin. Beuth
  - 108 DIN EN ISO 14971 (2001) Medizinprodukte - Anwendung des Risikomanagements auf Medizinprodukte. Berlin. Beuth
  - 109 DIN EN ISO 6385 (2004): Grundsätze der Ergonomie für die Gestaltung von Arbeitssystemen. Berlin. Beuth
  - 110 DIN EN ISO 7250 (2009): Wesentliche Maße des menschlichen Körpers für die technische Gestaltung. Berlin. Beuth
  - 111 DIN EN ISO 8402 (1995): Sammlung von Begriffsdefinitionen im Zusammenhang mit der Qualität. Berlin. Beuth
  - 112 DIN EN ISO 9241 (2002): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten. Berlin. Beuth
  - 113 DIN EN ISO 9241-11 (2006) Ergonomie der Mensch-System-Interaktion - Teil 11: Anforderungen an die Gebrauchstauglichkeit. Berlin. Beuth

- 114 DIN EN ISO 9241-110 (2008): Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung. Berlin. Beuth
- 115 DIN EN ISO 9241-12 (2000): Ergonomische Anforderungen für Bürotätigkeit mit Bildschirmgeräten - Teil 12: Informationsdarstellung Berlin. Beuth
- 116 DIN EN ISO 9241-6 (1999): Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 6: Leitsätze für die Arbeitsumgebung. Berlin. Beuth
- 117 Dix A., Finlay J., Abowd G., Beale R. (1999): Human-Computer Interaction. New York: Prentice Hall
- 118 Dix A., Finlay J., Abowd, G., Beale R. (1995): Mensch Maschine Methodik. München. Prentice Hall
- 119 Donchin Y., Gopher D., Olin M., Badihi Y., Biesky M., Sprung C. L., Pizov R., Cote S. (2003): A look into the nature and causes of human errors in the intensive care unit. In: Qual Saf Health Care. 12, S.143–148
- 120 Doubleday A., Ryan M., Springett M., Sutcliffe A. (1997): A comparison of usability techniques for evaluating design. In: Designing Interactive Systems (DIS '97) Conference Proceedings. New York: ACM Press. S. 101-110
- 121 Dräger J. (2007): Regeln für den Umgang mit Medizinprodukten In: Deutsches Ärzteblatt, Jg. 104 Heft 15, S.845-847
- 122 Driscoll M.P. (Eds.). (2004): Handbook of research on educational communications and technology (3rd ed.). Mahwah NJ: Lawrence Erlbaum Associates
- 123 Dumas J. (2002): User Based Evaluation. In: Jacko J., Sears A. (Hrsg.) The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications, Chapter 56. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., S.1093 - 1117
- 124 Dumas J. S., Redish J. C. (1999): A Practical Guide to Usability Testing. Bristol UK. Intellect Books
- 125 Dumas J.S., Fox J.E. (2008): Usability Testing: Current Practice and future directions. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application, New York. Taylor & Francis. S. 1129-1149
- 126 Dutt A., Johnson H., Johnson P. (1994): Evaluating evaluation methods. In: Cockton G., Draper S. W., Weir G. R. S. (Eds.): People and computers IX. Cambridge, UK: Cambridge University Press. S.109-121
- 127 Dzida W. (1983): Das IFIP-Modell für Benutzungsschnittstellen. In: Office Management, Sonderheft 31
- 128 Dzida W., Hofmann B., Freitag R., Redtenbacher W., Baggen R., Zurheiden C., Geis T., Beimel J., Hartwig R., Hampe-Neteler W., Peters H. (2000): Gebrauchstauglichkeit von Software. ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11. Bremerhaven: Wirtschaftsverl. NW (Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Forschungsbericht; Fb 921)
- 129 E DIN IEC 80001 (2007): Anwendung des Risikomanagements für IT-Netzwerke mit Medizinprodukten (Entwurf). Berlin. Beuth
- 130 Edmunds H. (1999): The Focus Group Research Handbook. NTC/Contemporary Publishing Group, Chicago
- 131 EG Richtlinie 90/385/EWG des Europäischen Parlaments und des Rates vom 20. Juni 1990 zur Angleichung der Rechtsvorschriften der Mitgliedstaaten über aktive implantierbare medizinische Geräte
- 132 EG Richtlinie 93/42/EWG des Europäischen Parlaments und des Rates vom 14. Juni 1993 über Medizinprodukte (Medizinprodukterichtlinie)

- 133 EG Richtlinie 98/79/EG des Europäischen Parlaments und des Rates vom 27. Oktober 1998 über In-vitro-Diagnostika
- 134 Ehrlich K., Rohn J. A. (1994): Cost Justification of Usability Engineering: A Vendor's Perspective. In: Bias R. G., Mayhew D. J., (Eds.) Cost-Justifying usability. San Francisco. Morgan Kaufmann Publishers. S.73-110
- 135 Eichstätt H., Schmidtke H. (1988): Physiologische und Psychologische Grenzen menschlichen Leistungsvermögens. In: Masing: Handbuch der Qualitätssicherung S. 725-750
- 136 Etchells E., Bailey C., Biason R., DeSousa S., Fowler L., Johnson K., Morash C., O'Neill C. (2006): Human Factors in Action: Getting "Pumped" at a Nursing Usability Laboratory. In: Healthcare Quarterly, 9(Sp) 2006. S. 69-74
- 137 Europäische Stiftung zur Verbesserung der Lebens- und Arbeitsbedingungen (2000). Zitiert in: Perlach E.: Vorwort zur Europäischen Woche für Sicherheit und Gesundheitsschutz bei der Arbeit 2000 – Belastung des Muskel- und Skelettsystems bei der Arbeit. Hauptverband der gewerblichen Genossenschaften (Hrsg.): S.11
- 138 Faulkner L. (2003): Beyond the five-user assumption. Benefits of increased sample sizes in usability testing. In: Behavior Research Methods, Instruments, & Computers 2003, 35 (3). S. 379-383
- 139 Felber A., Müller A. (2003): Sicherheit im Umgang mit Medizinprodukten. In: Gynäkologe 2004 · 37:45–50
- 140 Fields R., Paternò F., Santoro C., Tahmassebi S. (1999): Comparing Design Options for Allocating Communication Media in Cooperative Safety-Critical Contexts: A Method and a Case Study. In: ACM Transactions on Computer-Human Interaction, Vol. 6, No. 4. S. 3
- 141 Fredericks M., Basili V. (1998): Using Defect Tracking and Analysis to Improve Software Quality. Maryland. US Air Force Reserach Laboratory
- 142 Fu L., Salvendy G., Turley L. (2002): Effectiveness of user testing and heuristic evaluation as a function of performance classification. Behaviour & Information Technology, 21(2). S.137-143
- 143 Fu L., Schmidt K. (2006): Usability Evaluation. In: Kerwowski: International Encyclopedie of Ergonomics and Human Factors. S. 1154-1157
- 144 Gaba D.M., Fish K.J., Howard S.K. (1998): Zwischenfälle in der Anästhesie – Prävention und Management. Jena. Gustav-Fischer
- 145 Galitz W.O. (2007): The Essential Guide to User Interface Design. An Introduction to GUI Design Principles and Techniques. New York. Wiley, Third Edition
- 146 Galliers J., Sutcliffe A., Minocha A. (1999): An Impact Analysis Method for Safety-Critical User Interface Design. In: ACM Transactions on Computer-Human Interaction. Vol. 6, No. 4. S. 341-369
- 147 Gediga G., Hamborg K.-C. (2000): Evaluation in der Software-Ergonomie. Methoden und Modelle im Software-Entwicklungsprozess. In: Zeitschrift für Psychologie, 210 (1). S.40 - 57
- 148 Gediga G., Hamborg K.-C., Düntsch I. (1999):The IsoMetrics Usability Inventory: An operationalisation of ISO 9241-10. Behaviour and Information Technology, 18. S.151- 164
- 149 Geißler N. Wedekind M., Strauß G., Korb W. (2009): Analyse seltener Fehler bei der Mensch-Maschine-Interaktion im Operationssaal. In: i-com 1/2009 - Zeitschrift für interaktive und kooperative Medien. Oldenbourg Wissenschaftsverlag. S.38-42
- 150 Gerken J. (2002): Validität und Aussagekraft von Usability Test Methoden. In: Seminar zum Projektpraktikum. Universität Konstanz. FB Informatik und Informationswissenschaft; Arbeitsgruppe Mensch-Computer Interaktion
- 151 Gesellschaft Arbeit und Ergonomie - online e.V.: Ergo Online. Abruf unter: <http://www.ergo-online.de> (04.11.2008)

- 152 Gilbert J.E., Williams A., Seals C.D. (2007): Clustering for Usability Participant Selection. In: Journal of Usability Studies Vol. 3, Issue 1. S. 40-52
- 153 Glinz M. (2006): VL Modellierung: Kapitel 6 Interaktionsmodelle. Universität Zürich. Institut für Informatik, Abruf unter: [www.ifi.unizh.ch/groups/req/ftp/inf\\_II/kapitel\\_06.pdf](http://www.ifi.unizh.ch/groups/req/ftp/inf_II/kapitel_06.pdf) (30.05.2005)
- 154 Görner C. (2003): Styleguide - Vom Ladenhüter zum Steuerungsinstrument. In: Machate J., Burmeister M. (Hrsg.): User Interface Tuning – Benutzerschnittstellen menschlich gestalten. Frankfurt. Software und Support. S. 139-164
- 155 Görner C., Ilg R. (1993): Evaluation der Mensch-Rechner-Schnittstelle. In: Ziegler J., Ilg R. (Hrsg.): Benutzergerechte Software-Gestaltung. Standards, Methoden und Werkzeuge. München, Wien, Oldenbourg. S. 189-206
- 156 Gosbee J. (2002): Human factors engineering and patient safety. In: Qual. Saf. Health Care 2002,11,S.352-354
- 157 Graf J., Janssens U., Max M. (2007): Fehler und Zwischenfälle in der Intensivmedizin – Eine Begriffs- und Standortbestimmung In: Intensivmed. 44 S.119–128
- 158 Graham M., Kubose T., Jordan D., Zhang J., Johnson T., Patel V. (2004): Heuristic evaluation of infusion pumps: implications for patient safety in Intensive Care Units. In: International Journal of Medical Informatics, Volume 73, Issue 11, S. 771-779
- 159 Greenberg S. (2004): Working Through Task-Centered System Design. In: Diaper D., Stanton N. A.: The Handbook of Task Analysis for Human-Computer Interaction. Mahwah, New Jersey. Edited by Bournemouth University. Lawrence Erlbaum Associates. S.49-67
- 160 Hackos J. T., Redish J.C. (1998): User and task analysis for interface design. New York, NY. John Wiley & Sons
- 161 Hagen E., Mays G. (1981): Human Factors Engineering in the U.S. Nuclear Arena. In: Nuclear Safety, 17(3). S. 315-326. Zitiert in: Badke-Schaub P., Hofinger G., Lauche K.: Human Factors: Psychologie sicheren Handelns in Risikobranchen. Berlin, Heidelberg : Springer
- 162 Halbes S. (2003): Der vergleichende Warentest zur Unterstützung des nachhaltigen Konsums. Lehr- und Forschungsbericht Nr.52. Hannover. Universität Hannover
- 163 Hallbeck M.S., Koneczny S., Büchel D., Matern U. (2008): Ergonomic Usability Testing of Operating Room Devices. In.: Studies in health technology and informatics 132:147-52
- 164 Hamacher N., Kraiss K.F. (2004): Expertensystem zur kriterienorientierten Bewertung der Gebrauchsfähigkeit von Dialogsystemen. In: VDE-Kongress "Innovationen für Menschen" , Volume 1, Berlin. VDE Verlag GmbH. S. 193-198
- 165 Hamborg K.C., Hoemske T., Ollermann F. (2006): Qualitätssicherung im Usability Testing - zur Reliabilität eines Klassifikationssystems für Nutzungsprobleme. In: Heinecke A., Paul H. (Hrsg.): Mensch & Computer Mensch und Computer im Strukturwandel. München, Oldenbourg. S. 115-124
- 166 Hamborg K.-C., Ollermann F., Vehse B (2006): Zur Äquivalenz der Online- und Papier-Bleistift-Version des IsoMetrics Fragebogens zur Evaluation von Software. Zeitschrift für Arbeitswissenschaft, 60 S.123-132
- 167 Hansen W.J. (1971): User engineering principles for interactive systems. Proceedings of the Fall Joint Computer Conference 39. S. 523-532
- 168 Harloff J. (2004): Anforderungen an Usability-Verfahren – Relevanz für die Qualitätssicherung. In: GC UPA Arbeitspapier des AKs „Zertifizierung“ Anforderungen an Usability-Verfahren. UPA
- 169 Harper B.D., Norman K.L. (1993): Improving user satisfaction: The questionnaire for user interaction satisfaction version 5.5. In Proceedings of Mid Atlantic Human Factors Conference. Virginia Beach., S.224-228.

- 170 Hartmann B., Giemsa M., Gütschow S., Hanse H.-J., Hauck A. (2000): Das Bauspezifische Rückentraining für Auszubildende – Konzeption, Programm und Ergebnisse. Teil 1: Theoretische Grundlagen und Konzeption. In: ERGO-MED Zeitschrift für die arbeitsmedizinische und arbeitshygienische Praxis, 24, Heft 4, S. 168-174
- 171 Hassenzahl M., Prümper J., Buchbinder E. (1998): Software-ergonomische Beratung in der Praxis – ein Beitrag zur Organisationsentwicklung. In: Clermont A., Schmeisser W. (Hrsg.): Betriebliche Personal- und Sozialpolitik. München. Vahlen. S. 551-566
- 172 Haynes A.B., Weiser T.G., Berry W.R., et al. (2009): A surgical safety checklist to reduce morbidity and mortality in a global population. In: N Engl J Med 2009;360. S.491-499
- 173 Hegner M. (2003): Methoden zur Evaluation von Software. Informationszentrum Sozialwissenschaften, IZ-Arbeitsbericht Nr. 29. Abruf unter: [www.gesis.org/Publikationen/Berichte/IZ\\_Arbeitsberichte/pdf/ab\\_29.pdf](http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/pdf/ab_29.pdf) (10.07.2005)
- 174 Heidenberger B.: Checkliste Ergonomie am Arbeitsplatz. Abruf unter: <http://www.zeitblueten.com/files/Checkliste-Ergonomie.pdf> (20.02.2008)
- 175 Heinsen S. (2003): Das Usability-Lab. In: Heinsen S., Vogt P. (Hrsg.), Usability praktisch umsetzen. München: Hanser. S. 206-221
- 176 Helander M. G., Landauer T. K., Prabhu P. V. (Eds.) (1997): Handbook of human-computer interaction (2nd ed.). New York: North-Holland
- 177 Helbling H., Weinbeck T. (2004): Human-FMEA – Failure Mode and Effect Analysis. Diplomarbeit. Zürich. Abruf unter: <http://www.fmeainfocentre.com/foreign%20language/Human-FMEA.pdf> (25.11.2008)
- 178 Held J., Brüesch M., Zollinger A., Pasch T., Krueger H. (2002): Beteiligungsorientierte Arbeitsplatzanalyse. Schwachstellen im Arbeitssystem der Anästhesie in einem multidisziplinären OP-Trakt. In: Der Anaesthetist (2002), 51:S.110-115, Springer-Verlag
- 179 Helmreich RL. (2000): On error management: lessons from aviation. In: BMJ 320: S.781-785
- 180 Herczeg M. (2004): Interaktions- und Kommunikationsversagen in Mensch-Maschine-Systemen als Analyse- und Modellierungskonzept zur Verbesserung sicherheitskritischer Technologien. In: Grandt M.: Verlässlichkeit der Mensch-Maschine-Interaktion. DGLR-Bericht 2004-03. Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt
- 181 Hertzum M., Jacobsen N. E. (1999): The evaluator effect during first-time use of the cognitive walkthrough technique. In: Bullinger H.-J., Ziegler J. (Eds.): Human-Computer Interaction: Ergonomics and User Interfaces. Proceedings of the HCI International '99 (Vol. I, S. 1063-1067). London: Lawrence Erlbaum.
- 182 Hertzum M., Jacobsen N.E. (2001): The Evaluator Effect: A Chilling Fact about Usability Evaluation Methods. In: International Journal of Human-Computer Interaction, Vol. 13, No. 4 (2001). S. 421-443
- 183 Heuer J. (2003): Expertenevaluation. In: Heinsen S., Vogt P. (Hrsg.): Usability praktisch umsetzen – Handbuch für Software, Web, Mobile Devices und interaktive Produkte, Carl Hanser Verlag, München, Wien. S. 115-135
- 184 Hollingsed T., Novick D.G. (2007): Usability Inspection Methods after 15 Years of Research and Practice, In.: SIGDOC'07, October 22–24, 2007, El Paso. S. 249 - 255
- 185 Hollnagel, E. (1993): Human Reliability Analysis—Context and Control. Inc., New York, NY. Academic Press
- 186 Hölscher U., Laurig W. (2004): Ergonomie-Leitfaden für die Gestaltung von Medizinprodukten. Sankt Augustin. Kommission Arbeitsschutz und Normung Verein zur Förderung der Arbeitssicherheit in Europa. KAN-Bericht 31.

- 187 Hölscher U., Laurig W., Müller-Arnecke H. W. (2008):Prinziplösungen zur ergonomischen Gestaltung von Medizingeräten - Erkenntnisse, Empfehlungen und Prinziplösungen zur ergonomischen Produktgestaltung am Beispiel der Gebrauchstauglichkeit von Medizingeräten. Forschung Projekt F 1902. 2. Aufl.. Dortmund, Berlin. Bundesanst. für Arbeitsschutz und Arbeitsmedizin.
- 188 Holzinger A, Leitner H. (2005): Lessons from Real-Life Usability Engineering in Hospital: From Software Usability to Total Workplace Usability. In: Holzinger A., Weidmann K.-H. (Eds.) Empowering Software Quality: How can Usability Engineering reach these goals? Vienna, Austrian Computer Society, S. 153-160.
- 189 Holzinger A. (2001): Basiswissen Multimedia: Design. Würzburg. Vogel Buchverlag
- 190 Holzinger A. (2005): Usability Engineering Methods for software developers. In: Communications of the ACM, 48(1) S.71-74
- 191 Holzinger A. (2006): HCI4MED - Human-Computer Interaction - Usability Engineering in Medizin und Gesundheitswesen. In: Das Magazin der Österreichischen Computer Gesellschaft. Ausgabe 5/2006 Dezember - 31. Jhg S.24-25
- 192 Holzinger, A. (2006): Thinking-aloud - eine Königsmethode im Usability Engineering. OCG Journal 2006; 31(1). S. 4-5.
- 193 Howarth J.R. (2007): Supporting Novice Usability Practitioners with Usability Engineering Tools. Dissertation. Blacksburg. Virginia Tech
- 194 Hüttenrauch (1988): Gebrauchstauglichkeit. In: Masing W. Handbuch der Qualitätssicherung. München, Wien, Carl Hanser Verlag
- 195 Hvannberg E.T., Law E.L.-C. (2003): Classification of Usability Problems (CUP) Scheme, In: Interact (2003), S. 655-662.
- 196 Ilg R. (1993): Benutzerschnittstellen in der Normung. In: Ziegler, J., Ilg, R. (Hrsg.): Benutzergerechte Software-Gestaltung. Standards, Methoden und Werkzeuge. München, Wien, Oldenbourg S. 9-23
- 197 Isaksen U., Bowen J., Nissanke N . (1996): System and Software Safety in Critical Systems. Reading. University of Reading
- 198 ISO 26262: Road vehicles – Functional safety (in Entstehung)
- 199 ISO/IEC 25000 (2005): Software-Engineering - Qualitätskriterien und Bewertung von Softwareprodukten (SQuaRE) - Leitfaden für SQuaRE. Berlin. Beuth
- 200 ISO/IEC 25062 (2006): Software-Engineering - Qualitätskriterien und Bewertung von Softwareprodukten (SQuaRE) - Gemeinsames Industrieformat (CIF) für Berichte über Gebrauchstauglichkeitsprüfungen. Berlin. Beuth
- 201 Jacobs P. (1990): Intensivmedizinischer Arbeitsplatz aus Sicht der Pflege. In: Friesdorf W., Schwilk B., Hähnel J. (Hrsg.): Ergonomie in der Intensivmedizin. S.11-17
- 202 Janß A., Lauer W., Radermacher K.: Bewertung sicherheitskritischer Systeme im Operationssaal. In: i-com 1/2009 - Zeitschrift für interaktive und kooperative Medien. Oldenbourg Wissenschaftsverlag. S S.32-38
- 203 Jastrzebowski W. (1857): Grundriss der Ergonomie oder Wissenschaft von der Arbeit, gegründet auf der Wahrheit der Naturwissenschaft. In: Przyroda i Przemyst, Jg. 2, Nr. 29. S. 227-231
- 204 Jeffries R., Desurvire H. (1992) : Usability Testing vs. Heuristic Evaluation: Was there a contest? In: SIGCHI Bulletin. Volume 24, Number 4 S.39-41
- 205 Jeffries R., Miller J. R., Wharton C., Uyeda K. M. (1991): User interface evaluation in the real world. A comparison of four techniques. In CHI '91 Conference Proceedings. New York, ACM Press. S. 119-124



- 206 Johannson G. (1993): Mensch-Maschine-Systeme. Berlin, Heidelberg, New York. Springer-Verlag
- 207 Jonassen D.H., Hannum W.H., Tessmer M. (1989): Handbook of Task Analysis Procedures Westport: Greenwood Publishing Group
- 208 Jørgensen A. H. (1989): Using the thinking-aloud method in system development. In: Salvendy G., Smith M. J. (Hrsg.): Designing and using human-computer interfaces and knowledge based systems. Proceedings of HCI International 89, Volume 2. Advances in Human Factors/Ergonomics, Volume 12B. Amsterdam: Elsevier. S. 743– 750
- 209 Kage U. (2005): Risikosteuerung und Gefahrenabwehr im Medizinproduktegesetz. In: Veröffentlichungen des Instituts für Deutsches, Europäisches und Internationales Medizinrecht, Gesundheitsrecht und Bioethik der Universitäten Heidelberg und Mannheim: Das Medizinproduktegesetz – Staatliche Risikosteuerung unter Einfluss europäischer Harmonisierung. Volume 21. Springer Berlin Heidelberg S.203-289
- 210 Karat C., Campbell R. L., Fiegel T. (1992): Comparison of empirical testing and walk-through methods in user interface evaluation. In CHI '92 Conference Proceedings. New York: ACM Press. S. 397-404
- 211 Karat C.-M. (1994): A Business Approach to Usability Cost Justification. In: Bias R. G., Mayhew D. J.(Hrsg.): Cost-Justifying Usability, San Francisco. Academic Press. S. 45-70
- 212 Kaye R., Crowley J. (2000): Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management – Identifying, Understanding, and Addressing Use-Related Hazards. U.S. Department of Health and Human Services Food and Drug Administration. Abruf unter: <http://www.fda.gov/cdrh/humfac/1497.pdf> (10.07.2008)
- 213 Keenan S. L., Hartson H. R., Kafura D. G., Schulman R. S. (1999): The usability problem taxonomy: a framework for classification and analysis. In: Empirical Software Engineering, 4. S. 71 104
- 214 Kessner R. (2001): On the reliability of Usability testing. In: CHI 2001 Extended Abstracts. New York, NY: ACM, S. 97-98
- 215 Kies J. K., Williges R. C., Rosson M. B. (1998): Coordinating computer-supported cooperative work: A review of research issues and strategies. In: Journal of the American Society for Information Science, 49(9). S. 776-779
- 216 Kirakowski J., Corbett M. (1993): SUMI: The Software Usability Measurement Inventory. In: British Journal of Educational Technology Vol. 24, No. 3
- 217 Kirwan B., Aainsworth L.K. (1992): A Guide to Task Analysis. London. Taylor and Francis
- 218 Kitajima M., Polson P. G. (1992): A computational model of skilled use of graphical user interfaces. CHI '92 Proceedings: Human factors in computing systems, ACM, New York
- 219 Klauber J, Robra B.-P., Schellschmidt H. (2007): Schwerpunkt: Krankenhausvergütung - Ende der Konvergenzphase? In: Krankenhausreport 2007. Bonn. Schattauer
- 220 Kohn L.T., Corrigan J.M., Donaldson M.S. (1999): To err is human: building a safer health system. IOM-Report. Institute of Medicine. Washington DC. National Academy Press
- 221 Koneczny S. (2008): Erfassung und Analyse von Schwachstellen in der Funktionsstelle OP deutscher Krankenhäuser. Dissertation, Medizinische Fakultät Universität Tübingen
- 222 Kreuzig H.W. (1996): Wahrnehmungspsychologie. In: Dörner D, Herbert S. (Hrsg.): Psychologie - Eine Einführung in ihre Grundlagen und Anwendungsfelder. 2. Auflage. Stuttgart, Berlin, Köln. Verlag W. Kohlhammer
- 223 Krömker H. (2006):Usability - Stand der Forschung. In: Hennig J., Tjarks-Sobhani M. (Hrsg.):Usability und Technische Dokumentation. Lübeck. Schmidt-Römhild
- 224 Kunert T. (2009): User-Centered Interaction Design Patterns for Interactive Digital Television Applications. Human-Computer Interaction Series. Springer

- 
- 225 Kuniavsky M (2008): User experience and HCI. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application, New York. Taylor & Francis. S. 897-916
- 226 Kushniruka A., Triolab M., Steinc B., Boryckid E., Kannrye J. (2004): The Relationship of Usability to Medical Error: An Evaluation of Errors Associated with Usability Problems in the Use of a Handheld Application for Prescribing Medications. In: Stud Health Technol Inform. 2004;107(Pt 2): S. 1073-1076.
- 227 Lamsdale A., Chisholm S., Gagnon R., Davies J., Caird J. (2008): A Usability Evaluation of An Infusion Pump by Nurses Using a Patient Simulator. In: Human Factors and Ergonomics Society Annual Meeting Proceedings, Health Care. S. 1024-1028(5)
- 228 Lange W., Windel A. (2006): Kleine Ergonomische Datensammlung, 11. aktualisierte Ausgabe TÜV Media
- 229 Leape L.L. (1994): The Preventability of Medical Injury. In: M.S. Bogner (Hrsg.) Human Error in Medicine. Hillsdale NJ: Erlbaum Publ.. S.13-26
- 230 Leape L.L., Berwick D. M. (2005): Five Years After To Err Is Human - What Have We Learned? In: JAMA, May 18 2005, Vol 293, No. 19. S. 2384 - 2390
- 231 Leape L.L., Berwick D.M. (2000): Safe health care. Are we up to it? In: British Medical Journal, BMJ 2000; 320, BMJ Publishing Group Ltd S.725-726
- 232 Lenke N., Lutz H.-D., Sprenger M. (1995): Grundlagen sprachlicher Kommunikation. Mensch, Welt, Handeln, Sprache, Computer. München. UTB-Verlag
- 233 Lenorovitz D.R., Phillips M.D. (1987): Human factors requirements engineering for air traffic control system. In: Salvendy G., (Eds.): Handbook of human factors. New York: John Wiley. S.1771–89
- 234 Leveson N.G. (2001): Safeware – System safety and computers. Amsterdam. Addison-Wesley
- 235 Lewis C. (1982): Using the "thinking-aloud" method in cognitive interface design. In: IBM Research Report RC 9265. Yorktown Heights, NY: IBM T. J. Watson Research Center
- 236 Lewis C., Polson P., Wharton C., Rieman J. (1990): Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-And-Use Interfaces. Proceedings of the ACM CHI'90 Conference. Seattle: ACM. S. 235-241
- 237 Lewis C., Wharton C. (1997): Cognitive Walkthroughs. In: Helander M.G., Landauer T.K., Prabhu P. (Hrsg.): Handbook of Human-Computer Interaction. Amsterdam: Elsevier Science B.V. S.
- 238 Lewis J. R. (1982): Testing Small System Customer Setup. In: Proceedings of the Human Factors Society 26th Annual Meeting. S. 718-720
- 239 Lewis J. R. (1990): A psychometric evaluation of a post-study system usability questionnaire: The PSSUQ (IBM Tech. Report 54. 535). Boca Raton
- 240 Lewis J. R. (1994): Sample sizes for usability studies: Additional considerations. In: Human Factors, 36, 1994. S.368-378
- 241 Lewis J. R. (2001): Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. In: Int. J. Hum.-Comput. Interac, 13, S.445-480
- 242 Lewis J.R. (1991): Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. In: SIGCHI Bulletin Vol. 23 No.1 S.78-81
- 243 Lewis J.R. (1995): IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. In: Int. J. Hum.-Comput. Interact., Vol. 7, No. 1. S. 57-78.
- 244 Lewis J.R. (2002): Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. In: Int. J. Hum.-Comput. Interac, Vol. 14, Issue 3&4, 2002. S. 463-488

- 
- 245 Lewis J.R. (2006): Effect of Level of Problem Description on Problem Discovery Rates: Two Case Studies. In: Human Factors and Ergonomics Society Annual Meeting Proceedings, Test and Evaluation. S. 2567-2571
- 246 Lin L., Vicente K. J., Doyle D.J. (2001): Patient Safety, Potential Adverse Drug Events, and Medical Device Design: A Human Factors Engineering Approach. In: Journal of Biomedical Informatics 34. S. 274–284
- 247 Lowery J. C., James B. M. (1990): Evaluation of Health Care Software from a Usability Perspective. In: Journal of Medical Systems, Vol. 14, Nos. 1/2
- 248 Machate J., Burmeister M. (Hrsg.) (2003): User Interface Tuning – Benutzerschnittstellen menschlich gestalten. Software & Support Verlag GmbH
- 249 Mack R., Robinson J.B. (1992): When novices elicit knowledge: Question-asking in designing, evaluating, and learning to use software. In: Hoffman R.R.(Eds.), The psychology of expertise: Cognitive research and empirical AI. New York. Springer-Verlag. S.245-268.
- 250 Maguire M. (2001): Context of Use within usability activities. In: Int. J. Human-Computer Studies 55. S.453-483
- 251 Mahajan R., Shneiderman B. (1996): Visual & Textual Consistency Checking Tools for Graphical User Interfaces. In: IEEE Transactions on Software Engineering 23, 11 (November 1997). S. 722-735
- 252 Marcus A. (2002): Return on Investment for Usable User-Interface Design: Examples and Statistics. Abruf unter:  
[http://www.amanda.com/resources/ROI/AMA\\_ROIWhitePaper\\_28Feb02.pdf](http://www.amanda.com/resources/ROI/AMA_ROIWhitePaper_28Feb02.pdf) (04.07.2008)
- 253 Marsolek I. (2003): Entwicklung einer arbeitswissenschaftlichen Methodik zur Analyse und Optimierung von komplexen Prozessflüssen im Arbeitssystem Krankenhaus – TOPICS. Dissertation der Fakultät V der TU Berlin
- 254 Marsolek I., Friesdorf W. (2001): Prozesse optimieren – Umsetzung in der Klinik. In: Friesdorf W., Göbel M.: Effizienz steigern in OP und Intensiv. Berlin: TU-Press S. S.29-38
- 255 Matern U. (2004): Der Arbeitsplatz des Chirurgen: Untersuchungen zur Ergonomie und Verbesserungsvorschläge. Habilitationsschrift. Tübingen. Bisher unveröffentlicht
- 256 Matern U., Koneczny S., Scherrer M., Gerlings T. (2006): Arbeitsbedingungen und Sicherheit am Arbeitsplatz OP. In: Dtsch Arztebl 2006; 103(47): A 3187–92
- 257 Matern U., Büchel D., Scherrer M. (2009): Validierung medizinischer Geräte auch unter Berücksichtigung des Menschlichen Faktors. In: Hermeneit A., Steffen A., Stockhardt J. (Hrsg.): Medizinprodukte planen, entwickeln, realisieren (Begleit-CD), Köln, TÜV Media GmbH
- 258 Mauro C. (2002): Professional Usability Testing and Return on Investment as it Applies to User Interface Design for Web-Based Products and Services. Whitepaper, Mauro New Media
- 259 Mayhew D. J. (1999): The Usability Engineering Lifecycle. A Practitioner's Handbook for User Interface Design. San Francisco. Morgan Kaufmann
- 260 Mayhew D. J., Mantei M. (1994): A Basic Framework for Cost-Justifying Usability Engineering. In: Bias R. G., Mayhew D. J. (Hrsg.): Cost-Justifying Usability, San Francisco. Academic Press. S. 9-43.
- 261 Mayhew D.J. (2008): Requirement Specification within the usability engineering Lifecycle. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application New York. Taylor & Francis. S. 917-926
- 262 MEDDEV-Guidelines for the classification of medical devices (2001). Abruf unter:  
[http://ec.europa.eu/enterprise/medical\\_devices/meddev/2\\_2\\_4-1part2\\_07-2001.pdf](http://ec.europa.eu/enterprise/medical_devices/meddev/2_2_4-1part2_07-2001.pdf) (08.09.2009)

- 
- 263 Meister D. (1971): Human Factors. Theory and Practice. New York. John Wiley
- 264 Meister D. (1999): The history of human factors and ergonomics. Mahwah, New Jersey, London. Lawrence Erlbaum Associates
- 265 Merz J., Schütz A. (1996): Psychodiagnostik. In: Dörner D, Herbert S. (Hrsg.): Psychologie - Eine Einführung in ihre Grundlagen und Anwendungsfelder 2. Auflage. Stuttgart. Kohlhammer. S. 231-248
- 266 Meyer B., Skiba D. (2007): Usability-Testing Healthcare Software with Nursing Informatics Students in Distance Education: A case study. In: Holzinger A. (Eds.) HCI and Usability for Medicine and Health Care. USAB LNCS (2007) S.447-452
- 267 MIL-STD-1629A (1980): Procedures for performing a Failure Mode, Effects, and Criticality Analysis. Military Standard, US-Department of Defense, Washington DC, 1980
- 268 Molich R., Bevan N., Curson I., Butler S., Kindlund E., Miller D., Kirakowski J. (1998) : Comparative evaluation of usability tests. In: Proceedings of the Usability Professionals Association 1998 Conference, Chicago, IL: UPA.S. 189-200
- 269 Molich R., Dumas J. (2008): Comparative usability evaluation (CUE-4). In: Behaviour and Information Technology Volume 27, Issue 3, May 2008, S. 263-281
- 270 Molich R., Ede M., Kaasgraad K., Karyukin B. (2004): Comparative usability evaluation. In: Behaviour & Information Technology, vol. 23, no. 1. S. 65-74
- 271 Molich R., Nielsen J. (1990). Improving a human-computer dialogue. In: Communications of the ACM 33, 3. S. 338-348
- 272 Molich R., Thomsen A.D., Karyukina B., Schmidt L., Ede M., van Oel W., Arcuri M. (1999): Comparative evaluation of usability tests. In: Extended Abstracts of ACM CHI 99 Conference. New York: ACM Press.
- 273 Montag K. (2009): Untersuchung zur Gebrauchstauglichkeit von Medizinprodukten im Anwendungsbereich OP. Dissertation an der Medizinischen Fakultät, Universität Tübingen
- 274 Montag K., Rölleke T., Büchel D., Matern U. (2008): Untersuchung zur Gebrauchstauglichkeit von Medizinprodukten im Anwendungsbereich OP, Vortrag AHFE2008, Las Vegas
- 275 MPBetreibV. (2002) : Medizinprodukte-Betreiberverordnung. Fassung vom 21. August 2002. BGBl. I (2002) S. 3396
- 276 MPG. (2002): Medizinproduktegesetz. Fassung vom 07. August 2002. BGBl. I (2002) S. 3146
- 277 MPSV - Medizinprodukte-Sicherheitsplanverordnung. Verordnung über die Erfassung, Bewertung und Abwehr von Risiken bei Medizinprodukten. Ausfertigungsdatum: 24.06.2002
- 278 Müller C., Müller-Peters H., Scholzen J., Kiell G., Melles T., Maas J., Höllger T. (2004): Marktforschungs- Almanach. Das Taschenlexikon der Marktforschung, Psychonomics AG
- 279 Muller J.M., Matheson L., Page C., Gallup R. (1998): Methods & tools: participatory heuristic evaluation. In: Interactions Volume 5, Issue 5 (Sept./Oct. 1998). S. 13 - 18
- 280 Mutschler B., Reichert M. (2004): Usability-Metriken als Nachweis der Wirtschaftlichkeit von Verbesserungen der Mensch-Maschine-Schnittstelle. In: Proc. IWSM / MetriKon Workshop on Software Metrics (IWSM / MetriKon'04), Königs Wusterhausen
- 281 National Coordinating Council for Medication Error Reporting and Prevention (NCC MERP) (2001): Taxonomy of Medication Errors. Abruf unter: <http://www.nccmerp.org/pdf/taxo2001-07-31.pdf> (27.03.2009)
- 282 Nielsen J. (1992): Finding usability problems through heuristic evaluation. In: CHI '92 Conference Proceedings. New York: ACM Press. S. 373-380

- 283 Nielsen J. (1992): Reliability of severity estimates for usability problems found by heuristic evaluation. In: Conference on Human Factors in Computing Systems. Monterey. S. 129-130
- 284 Nielsen J. (1993): Usability Engineering. San Diego. Academic Press
- 285 Nielsen J. (1995): Technology Transfer of Heuristic Evaluation and Usability Inspection. International Conference on Human-Computer Interaction. Lillehammer, Norway
- 286 Nielsen J. (2005): Medical Usability: How to Kill Patients Through Bad Design. In: Jakob Nielsen's Alertbox, April 11. Abruf unter: [www.useit.com/alertbox/20050411.html](http://www.useit.com/alertbox/20050411.html) (20.12.2007)
- 287 Nielsen J. (2006): Usability Testing. In: G. Salvendy (eds.), Handbook of Human Factors and Ergonomics, New York. John Wiley (S.1543-1568)
- 288 Nielsen J., Landauer T. K. (1993): A mathematical model of the finding of usability problems. Conference on Human Factors in Computer Systems: CHI99. S. 83-84
- 289 Nielsen J., Mack R. I. (Hrsg.) (1994): Usability inspection method. New York. John Wiley
- 290 Nielsen J., Molich R. (1990): Heuristic evaluation of user interfaces. In: Proc. ACM CHZ'90 Seattle, WA. S. 249–256
- 291 Nolan D.(2008): Safety and Security Review for the Process Industries. 2nd Edition: Application of HAZOP. PHA and What-If Reviews. Elsevier Science & Technology
- 292 Norman D. (2002): The Design of Everyday Things, New York.: Doubleday
- 293 Norman D. A. (1981): Categorization of action slips. In: Psychological Review, 88(1), S.1-15.
- 294 Norman D. A. (1983): Design Rules Based on Analyses of Human Error. In: Communications of the ACM, Vol. 26, No. 4, April. S. 254–258
- 295 Norman D.A. (1986): Cognitive Engineering. In: Norman D.A., Draper S.W. (Eds.): User Centered System Design. Hillsdale: Lawrence Erlbaum Associates. S. 31-61
- 296 Nörr M.: Merkblatt (2006): CE-Kennzeichnung von Medizinprodukten. Abruf unter: <http://www.schwaben.ihk.de/dokumente/merkblaetter/M166819.pdf> (06.09.2008)
- 297 O'Hare D., Wiggins M., Batt R, Morrison D. (1994): Cognitive failure analysis for aircraft accident investigation. In: Ergonomics, 37, S.1855-1869.
- 298 Oestereich B. (2001): Objektorientierte Softwareentwicklung – Analyse und Design mit der Modeling Language UML. Oldenbourg, München
- 299 Ollenschläger G., Thomeczek C. (2002): Qualitätsmanagement im Gesundheitswesen: Fehlerprävention und Umgang mit Fehlern in der Medizin. In: Medizinische Klinik 2002; 97:564–70 (Nr. 9) München, Urban & Vogel,
- 300 Palanque P., Paternò F., Wright P. (1998): Designing user interfaces for safety critical systems. In: CHI 98. SIGCHI1998:
- 301 Park S.K. (1997): Human Error. In: Salvendy G.: Handbook of Human Factors and Ergonomics. New York. John Wiley & Sons
- 302 Paternò F. (2001): Task Models in Interactive Software Systems. In: Jonassen D., Hannum W., Tessmer M.: Handbook of Software & Knowledge Engineering. Westpoint World Scientific Publishing Co.. S. 192-206
- 303 Paternò F. (2004): ConcurTaskTrees: An Engineered Notation for Task Models. In: Diaper D., Stanton, N. A.: The Handbook of Task Analysis for Human-Computer Interaction. Mahwah, New Jersey. Edited by Bournemouth University. Lawrence Erlbaum Associates. S.483-503

- 
- 304 Paternò F., Santoro C., Fields R.E. (1999): Analysing user deviations in interactive safety-critical applications. In: Duke D.J., Puerta A. (Eds.): Proceedings of the Conference on Design, Specification, and Verification of Interactive Systems (DSV-IS '99). S.189–204
- 305 Patterson P. A. North R.A. (2006): Fitting Human Factors in the Product Development Process. Excerpt from Medical Device & Diagnostic Industry, January 2006. Abruf unter: [www.agilisconsulting.com/pdf/HumanFactorsandtheProductDevelopmentProcess.pdf](http://www.agilisconsulting.com/pdf/HumanFactorsandtheProductDevelopmentProcess.pdf) (10.08.2009)
- 306 Patterson P.A., North R.A. (2007): In Plain Sight, The Benefits of Human Factors Analysis. Regulatory Affairs Focus, January 2007 12(1), 12-15. Abruf unter: <http://www.humancenteredstrategies.com/resources.php> (10.07.2008)
- 307 Peeck J. (2000): Einführungsvortrag Rechner-Sicherheit und Firewalls. Universität Heidelberg 2000. Abruf unter: <http://web.urz.uni-heidelberg.de/Netzdienste/nm/Netzfort001212/sld001.htm> (14.05.2009)
- 308 Perrow C. (1999): Normal Accidents. Princeton. Princeton University Books
- 309 Peters O., Meyna A. (1988): Sicherheitstechnik. In: Masing W. (Hrsg.): Handbuch der Qualitätssicherung. München, Wien, Carl Hanser Verlag. S. 301-329
- 310 Petrick K., Reihlen H. (1988): Qualitätssicherung und Normung. In.: Masing W.: Handbuch der Qualitätssicherung. München, Wien, Carl Hanser Verlag S.51-63
- 311 Pierotti D. (1993): Usability Techniques: Heuristic Evaluation – A System Checklist. Abruf unter: <http://www.stcsig.org/usability/topics/articles/he-checklist.html> (30.12.2007)
- 312 Polson P.G., Lewis C., Rieman J., Wharton C. (1992): Cognitive Walkthroughs – A Method for Theory-Based Evaluation of User Interfaces. In: International Journal of Man-Machine-Studies 36. S. 741-773
- 313 Preece J., Rogers Y., Sharp H. (2002): Interaction Design: Beyond Human-Computer Interaction. New York, NY. John Wiley & Sons.
- 314 Preece J., Rogers Y., Sharp H., Benyon D., Holland S., Carey T. (Hrsg.) (1994): Human-Computer Interaction. Wokingham (UK), Addison-Wesley
- 315 Preim B. (1999): Entwicklung Interaktiver Systeme. Grundlagen, Fallbeispiele und innovative Anwendungsfelder. Berlin. Springer
- 316 Pressman R.S. (1992): Software Engineering: A Practitioner's Approach. New York. McGraw-Hill
- 317 Prümper J. (1999): TestIT: ISONORM 9241/10. In: H.-J. Bullinger & J. Ziegler (Eds.) Human-Computer Interaction. Ergonomics and User Interfaces. Vol. 1, Mahwah, NJ, Erlbaum. S. 1027-1028
- 318 Prümper J., Anft M. (1993): Die Evaluation von Software auf Grundlage des Entwurfs zur internationalen Ergonomie-Norm ISO 9241 Teil 10 als Beitrag zur partizipativen Systemgestaltung - Ein Fallbeispiel. In: Rödiger K-H. (Hrsg.): Software-Ergonomie '93, Teubner, Stuttgart. S.145-156.
- 319 Purho V.: Heuristic Inspections for Documentation – 10 Recommended Documentation Heuristics. In: STC Usability SIG Newsletter April 2000, Vol. 6, No. 4 - Usability and Documentation, Abruf unter: <http://www.stcsig.org/usability/newsletter/0004-docsheuristics.html> (26.06.2008)
- 320 Quesenbery W. (2001): Building a Better Style Guide. EServer Technical Communication Library. Abruf unter: <http://www.wqusability.com/articles/better-style-guide-paper.pdf> (22.01.2008)
- 321 Radermacher K. (1999): Computergestützte Operationsplanung und -ausführung mittels individueller Bearbeitungsschablonen in der Orthopädie. In: Rau (Hrsg.) Berichte aus der Biomedizintechnik. Aachen

- 
- 322 Rall M., Manser T., Guggenberger H., Gaba D.M., Unertl K. (2001): Patientensicherheit und Fehler in der Medizin. Entstehung, Prävention und Analyse von Zwischenfällen. In: *Anästhesiologie Intensivmedizin Notfallmedizin Schmerztherapie* (2001) 36:321–30
- 323 Raskin J. (2000): *The Humane Interface*. Reading, MA: Addison-Wesley Longman, Inc
- 324 Raskin J. (2001): *Das intelligente Interface*. München. Addison-Wesley
- 325 Rasmussen J. (1982): Human Errors. A Taxonomie for describing human malfunction in industrial installations. In: *Journal of occupational Accidents*, 4, 31. S.311-333
- 326 Rasmussen J. (1986): *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering*. New York. Elsevier
- 327 Rasmussen J. (1987): The Definition of Human Error and a Taxonomy for Technical System Design. In: Rasmussen J., Duncan K., Leplat J.: *New Technology and Human Error*. S.23-39
- 328 Rau G. (2008): Ergonomie in der Medizin und Berührungseingabe über Farbgraphik-Displays. In: Schmidt I., Schlick C.M., Grosche J. (Hrsg.) *Ergonomie und Mensch-Maschine-Systeme*. Springer
- 329 Reason J. (1987): A Preliminary Classification of Mistakes. In: Rasmussen J., Duncan K., Leplat J. *New Technology and Human Error*. Chichester. Wiley
- 330 Reason J. T. (1994): *Menschliches Versagen. Psychologische Risikofaktoren und moderne Technologien*. Heidelberg: Spektrum Akad. Verlag.
- 331 Reason J.(1987): Generic Error-Modeling System GEMS: A Cognitive Framework for Locating Common Human Error Forms In: Rasmussen J., Duncan K., Leplat J.: *New Technology and Human Error*. New Technology and Human Error. Chichester, New York. Wiley. S.63-83
- 332 Reason J.T. (1990): *Human error*. Cambridge. Cambridge University Press
- 333 REFA (Hrsg.) (1993): *Methodenlehre der Betriebsorganisation. Teil: Grundlagen der Arbeitsgestaltung*. REFA-Verband für Arbeitsgestaltung, Betriebsorganisation und Unternehmensentwicklung e. V. München, Hanser Verlag
- 334 Rice V. J. (2006): Medical Equipment Usability Testing. An Introduction. In: W. Karwowski (Eds.) *International Encyclopedia of Ergonomics and Human Factors*. Boca Ration. CRC Press S.3243 - 3248
- 335 Richard A., Stephans J. (2000): *System Safety for the 21st Century. The Updated and Revised Edition of System Safety 2000*. Hobaken, New Jersey. Wiley-Interscience
- 336 Richter M., Flückiger M. (2007): *Usability Engineering kompakt. Benutzbare Software gezielt entwickeln*. München. Spektrum Akademischer Verlag.
- 337 Rieser S. (2001): Bewältigung des demographischen Wandels. Mehr Wettbewerb und mehr Prävention als Rezept. In: *Deutsches Ärzteblatt* 98 B, 2001. S.189-190
- 338 Riihiahho S. (2002): The Pluralistic Usability Walk-Through. In: *Ergonomics in Design: The Quarterly of Human Factors Applications*, Volume 10, Number 3, 2002. Human Factors and Ergonomics Society. S. 23-27(5)
- 339 Ritter F., Hansen C., Wilkens K., Köhn A., Peitgen H.-O. (2009): Benutzerschnittstellen für den direkten Zugriff auf 3D-Planungsdaten im OP. In: *i-com 1/2009 - Zeitschrift für interaktive und kooperative Medien*. Oldenbourg Wissenschaftsverlag. S.24-31
- 340 Rogers M.L., Patterson E.S., Chapman R.J., Render M.L.. (2005): Usability Testing and the Relation of Clinical Information Systems to Patient Safety. In: Henriksen K., Battles J.B., Marks E., Lewin D.I.: *Advances in Patient Safety: from research to implementation*. Volume 2, AHRQ Publication Nos. 050021 (1-4). February 2005. Rockville, MD. Agency for Healthcare Research and Quality. S. 365-378.

- 341 Rögner A. (2005): Untersuchungen zur Funktion von Benutzerinformationen für die Beeinflussung der menschlichen Zuverlässigkeit in sozio-technischen Systemen. Dissertation: Lehrstuhl für Arbeitswissenschaften / BTU Cottbus
- 342 Rohmert W. (1978): Der Mensch als Einflussgröße auf die Zuverlässigkeit technischer Systeme. In: Qualität und Zuverlässigkeit 23 2. S. 44-51
- 343 Rohmert W., Haider E. (1988): Ergonomie der (Prüf-)Arbeitsplätze. In: Masing W. (Hrsg.) Handbuch der Qualitätssicherung. München, Wien, Carl Hanser Verlag. S. 751-770
- 344 Rölleke T. (2007): Misslungenes Design. Usability-Mängel als Gefährdung der Sicherheit. Vortrag auf dem World of Usability Day 2007. München
- 345 Rosson M. B., Carrol J. M. (2003): Scenario-based design. In: Jacko J.A., Sears A.: (Hrsg.): The Human computer interaction handbook – Fundamentals, Evolving Technologies and Emerging Applications. Mahwah, New Jersey. Lawrence Erlbaum Associates. Inc. Publisher. S. 1032–1050
- 346 Rowland K, Pozzo D. (2007): What is Human Factors Analysis? In: Science & Technology 2007. S.36-39
- 347 RTCA/DO-178B (1985): Software Considerations in Airborne Systems and Equipment Certification; EURCAEED-12B. Washington: Radio Technical Commission of Aeronautics. Paris: European Organisation for Civil Aviations Electronics
- 348 Rubin J. (1994): Handbook of Usability Testing. New York. John Wiley & Sons
- 349 Rudolf C. (2006): Handbuch Software-Ergonomie – Usability Engineering. Unfallkasse Post und Telekom. Abruf unter: <http://www.ukpt.de/pages/dateien/software-ergonomie.pdf> (29.07.2008)
- 350 Sarodnick F.; Brau H. (2006): Methoden der Usability Evaluation. Bern. Hubner-Verlag
- 351 Sauro J., Kindlund E. (2005): How long Should a Task Take? Identifying Specification Limits for Task Times in Usability Tests. In: Proceeding of the Human Computer Interaction International Conference CHI2005, Las Vegas
- 352 Sawyer D. (Eds.)(1996): Do it by Design – An Introduction to Human Factors in Medical Design. U.S. Department of Health and Human Services. Public Health Services. Food and Drug Administration. Center of Device and Radiological Health.
- 353 Schäppi B., Andreasen M.M., Kirchgeorg M., Radermacher F.-J. (2005): Handbuch Produktentwicklung. München, Wien. Hanser Fachbuchverlag
- 354 Schaub H. (2006): Störungen und Fehler beim Denken und Problemlösen. In: J. Funke (Eds.). Denken und Problemlösen. Enzyklopädie der Psychologie, Themenbereich C: Theorie und Forschung, Serie II: Kognition, Band 8. Göttingen. Hogrefe
- 355 Schaub H.: Menschliches Versagen. Die Rolle des Faktors "Mensch" bei großtechnischen Katastrophen aus psychologischer Sicht. Memorandum Nr. 35. Institut für theoretische Psychologie. Otto-Friedrich-Universität. Abrufbar unter: [www.daec.de/flusi/flytop/downfiles/schaub.pdf](http://www.daec.de/flusi/flytop/downfiles/schaub.pdf) (01.04.2009)
- 356 Scherhorn G. (2001): Warentest. In: Diller H. (Hrsg.): Vahlens Großes Marketinglexikon, München, S. 1840. Zitiert in: Halbes S.: Der vergleichende Warentest zur Unterstützung des nachhaltigen Konsums. Lehr- und Forschungsbericht Nr.52, Universität Hannover (2003). S. 1840-1841
- 357 Schiendzielorz E.(2008): Gebrauchstauglichkeitsstudie einer Spritzenpumpe. Masterarbeit in Kooperation mit der FH Schwäbisch Gmünd und dem Experimental-OP des Universitäts-Klinikums Tübingen
- 358 Schmidtke H: Ergonomische Bewertung von Arbeitssystemen. Entwurf eines Verfahrens. München. Hanser



- 359 Schrappe M., Lessing C., Albers B., Conen D., Gerlach F., Grandt D., Hart D., Jonitz G., Lauterberg J., Loskill H., Rothmund M.(2007): Agenda Patientensicherheit 2007. Aktionsbündnis Patientensicherheit e.V.. Witten
- 360 Schrenk C. (2004): Medizinproduktegesetz und Arzthaftung. In: di Fabio U., Vieweg K.: Reihe Recht - Technik - Wirtschaft, Band 94, Carl Heymanns Verlag
- 361 Seghezzi D. (1988): Qualitätssicherung von Neuprodukten. In: Masing W.: Handbuch der Qualitätssicherung. München, Wien, Carl Hanser Verlag. S. 331-359
- 362 Seitz T. (2007): Ergonomie in der Medizintechnik – Anforderungen an die Entwicklung gebrauchstauglicher Bedienkonzepte. In: mt Medizintechnik 127 (2007), Nr. 4, S. 133-135
- 363 Seitz T.: Gebrauchstauglichkeit von Medizinprodukten. Die Ergänzungsnorm DIN EN 60601-1-6. VDE. Abruf unter: <http://www1.vde.com/NR/rdonlyres/B9CA24E9-8777-41DF-83AB-EF417FC6C174/20679/IS545D.pdf> (08.07.2007)
- 364 Sexton J.B., Thomas E. J., Helmreich R.L. (2000): Error, stress, and teamwork in medicine and aviation: Cross sectional surveys. In: BMJ 2000;320. S. 745-749
- 365 Shackel B. (1991): Usability – context, framework, design and evaluation. In: Shackel B., Richardson S. (Eds.): Human Factors for Informatics Usability. Cambridge. Cambridge University Press. S.21-38
- 366 Shalala D., Herman A., Eisenberg J. (2000): Doing What Counts for Patient Safety. Federal Actions to Reduce Medical Errors and Their Impact. Report of the Quality Interagency Coordination Task Force (QuIC). Abruf unter: [www.quic.gov/report/errors6.pdf](http://www.quic.gov/report/errors6.pdf) (06.07.2009)
- 367 Shannon C.E., Weaver W. (1949): The Mathematical Theory of Communication. Urbana. University of Illinois Press
- 368 Shneiderman B. (1992): Designing the User Interface. Strategies for Effective Human-Computer Interaction. Massachusetts. Addison-Wesely Publishing Company.
- 369 Shneiderman B. (1999): Das Bewusstsein wächst- Das goldene Zeitalter des Benutzers. In: SAP INFO Nr. 59 – 3/99
- 370 Shneiderman B. (2002): Leonardos Laptop: Human needs and the new computing technologies, Cambridge, Mass.: MIT Press
- 371 Shneiderman B.(2002): User Interface Design. (dt. Ausgabe) Bonn. mitp-Verlag,
- 372 Simeral E. J., Branaghan R. J. (1997): A Comparative Analysis of Heuristic and Usability Evaluation Methods. Proceedings of Society for Technical Communication 44th Annual Conference. S.307-308
- 373 Sommerville I.(2008): Software Engineering. 8. aktualisierte Auflage. München. Addison Wesley
- 374 Stapelkamp T. (2007): Screen- und Interfacedesign: Gestaltung und Usability für Hard- und Software. Berlin. Springer
- 375 Statistisches Bundesamt
- 376 Steinhauser (2002): Umsetzung der MPBetreibV im UKT. Vortrag. Steinhauser
- 377 Stempfle C.T. (2006): Der Arbeitsplatz OP aus Sicht des Juristen – Gesetze, die keiner (so) kennt. Vortrag auf der MEDICA Vision. Düsseldorf 17.11.2006
- 378 Stiftung Warentest (2004): Satzung, §2. In: Stiftung Warentest. 40 Jahre Stiftung Warentest
- 379 Stoessel S. (2004): Methoden des Testings im Usability Engineering. In: Beier. M., von Gizycki V. (Hrsg.): Usability – nutzerfreundliches Web-Design. Berlin. Springer. S. 75-96
- 380 Storey N. (1996): Safety critical computer systems. Harlow, England. Prentice Hall,
- 381 Störfallkommission beim Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (2004): Risikomanagement im Rahmen der Störfall-Verordnung. Bonn

- 382 Szalma J.L.; Hancock P.A. (2008): Task Loading and Stress in Human-Computer Interaction: Theoretical Frameworks and Mitigation Strategies. In: Sears A., Jacko J.A.: The Human Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Application, New York. Taylor & Francis. S.115-132
- 383 Thaller G.E. (2002): Interface Design. Die Mensch-Maschine-Schnittstelle gestalten. Konzepte für Programm- und Web-Oberfläche. Frankfurt: Software & Support Verlag.
- 384 The National Patient Safety Foundation at the AMA (1999): Agenda for Research and Development in Patient Safety. Abruf unter: <http://www.npsf.org/pdf/r/researchagenda.pdf> (04.12.2008)
- 385 Thimbleby H. (2007): Interaction Walkthrough: Evaluation of safety critical interactive systems. In: Doherty G., Blandford A. (Eds.): DSVIS 2006. LNCS, vol. 4323, Heidelberg. Springer. S. 52-66.
- 386 Thimbleby H. (2007): User-centered Methods are Insufficient for Safety Critical Systems. n: Holzinger A (Eds.): HCI and Usability for Medicine and Health Care. USAB 2007, LNCS. S. 1-20
- 387 Thomas C., Bevan N. (1997) (Eds.): Usability Context Analysis. A practical guide. Version 4. Teddington. NPL Usability Services
- 388 Thomeczek C, Bock W, Conen D, et al. (2005): Das Glossar Patientensicherheit – Ein Beitrag zur Definitionsbestimmung und zum Verständnis der Thematik „Patientensicherheit“ und „Fehler in der Medizin“. Gesundheitswesen. 2004 Dec; 66(12): 833-840; Ärztliches Zentrum für Qualität in der Medizin
- 389 Thyvalikakath T.P., Schleyer T.K., Monaco V. (2007): Heuristic evaluation of clinical functions in four practice management systems: a pilot study. In: Journal of the American Dental Association. 2007 Feb;138(2): S. 209-10
- 390 Tilley A.R: The Measure of Man and Woman: Human Factors in Design. New York. Henry Dreyfuss Associates
- 391 Timpe K.P., Jürgensohn T., Kolrep H. (2002): Mensch-Maschine-Systemtechnik. Konzepte, Modellierung, Gestaltung, Evaluation. Düsseldorf. Symposium Publishing GmbH
- 392 Tognazzini B. (2003): First Principles of Interaction Design. Nielsen Norman Group, 2003, Abruf unter: <http://www.asktog.com/basics/firstPrinciples.html>. (24.11.2007)
- 393 Troidl H. (2000): Risikoanalyse in der Chirurgie. In: Der Chirurg S.771-783
- 394 Turner C.W., Lewis J. R., Nielsen J. (2006) : Determining Usability Test Sample Size. In: International Encyclopedia of Ergonomics and Human Factors, Second Edition, Volume 3. Edited by Waldemar Karwowski, Boca Raton, FL: CRC Press
- 395 U.S. Department of Health & Human Services: How can I show that usability engineering saves money? Abruf unter: <http://usability.gov/basics/usasaves.html> (14.05.2009)
- 396 U.S. Department of Health and Human Services. Food and Drug Administration. Task Force on Risk Management (1999): Managing the Risks from Medical Product Use - Creating a Risk Management Framework
- 397 U.S. Food and Drug Administration (FDA) (2009): Code of Federal Regulations, Title 21, Vol. 8. Revised as of April 1, 2007] [CITE: 21CFR820.30] Abruf unter: [www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?FR=820.30](http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?FR=820.30) (1.1.2007)
- 398 Ulrich E. (2003): Arbeitspsychologie in Krankenhaus und Arztpraxis. Arbeitsbedingungen, Belastungen, Ressourcen. Reihe: Schriften zur Arbeitspsychologie 61. Bern. Huber
- 399 Ulrich E. (2005): Arbeitspsychologie. Auflage: 6, Stuttgart. Schäffer-Poeschel
- 400 Usability Professionals' Association: Resources: Usability in the Real World. Business Benefits. Bloomingdale. Abruf unter: [www.upassoc.org](http://www.upassoc.org) (18.05.2009)

- 
- 401 van Veelen M.A. (2003): Human-Product interaction in minimal invasive surgery: A Design Vision for Innovative Products. Dissertation. University of Technology Delft
- 402 van Veelen M.A., Meijer D.W., Goossens R.H.M., Snijders C.J., Jakimowicz J.J. (2002): Improved usability of a new handle design for laparoscopic dissection forceps. In: Surg Endosc (2002) 16: S. 201–207
- 403 VDI/VDE: 3850 Blatt 1-3 (2002): Nutzergerechte Gestaltung von Bediensystemen für Maschinen - Interaktionsgeräte für Bildschirme. Berlin. Beuth
- 404 Vincent C., Neale G., Woloshynowych M. (2001): Adverse events in British hospitals: preliminary retrospective record review. In: BMJ 322 (7285), 2001, S. 517-519.
- 405 Von der Mosel H.A. (1971): Der klinisch – biomedizinische Ingenieur. In: Schweizerische Ärztezeitung 52 vom 29. 12. 1971. Zitiert in: Matern U, Koneczny S, Scherrer M, Gerlings T.: Arbeitsbedingungen und Sicherheit am Arbeitsplatz OP. Dtsch Arztebl 2006; 103(47): A 31
- 406 Ward J., Clarkson J. (2007): Human Factors engineering and the design of medical devices. In: Carayon P.: Handbook of Human Factors and Ergonomics in Health Care and Patient Safety. Hillsdale, NJ. Lawrence Erlbaum Associates.
- 407 Watson J.A.(1986): Fundamental Constraints on Some Event Data. In: Proceedings of the 5th EuReData Conference. Heidelberg. Springer Verlag
- 408 Weinger M.B.: A Clinician's Perspective on Designing Better Medical Devices. In: Wiklund M., Wilcox S.: Designing Usability into Medical Products, Boca Ration. CRC Press. Foreword
- 409 Weller G., Schlag B. (2002): Kriterien zur Beurteilung von Fahrerassistenzsystemen. BDP-Kongress für Verkehrspsychologie. Regensburg.
- 410 Wharton C., Rieman J., Lewis C., Polson P. (1994): The cognitive walkthrough method. A practitioner's guide. In: Nielsen J., Mack R.L. (Eds.): Usability Inspection Methods. New York. John Wiley. S. 105-140
- 411 Whiteside J., Bennett J. Holtzblatt K. (1988): Usability Engineering. Our Experience and Evolution. In: Helander M.: The Handbook of Human Computer Interaction. Amsterdam. Elsevier Science Publishers. S. 791-817
- 412 Wiegmann D. A., Shapell S. A. (1997): Human factors analysis of postaccident data: Applying theoretical taxonomies of human error. In: The International Journal of Aviation Psychology, 7 (1). Lawrence Erlbaum Associates, Inc. S.67-81
- 413 Wiesollek D., Müller-Arnecke H. W., Hold U. (2008): Untersuchungen zur Ergonomie von Medizinprodukten – Fallbeispiel Spritzenpumpen. Dortmund. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin
- 414 Wiklund M. (1995): Medical Device and Equipment Design – Usability Engineering and Equipment Design. Boca Ration. Buffalo Grove. Interpharm Press.
- 415 Wiklund M. (2000): Human Factors Making Medical Device Interfaces More User-Friendly. In: Wiklund M., Wilcox S.: Designing Usability into Medical Products. Boca Ration. CRC Press. S.151-160
- 416 Wiklund M.E., Carstensen P.B., Hasler R.A., Sawyer D., Weinger M.B. (2001): Human Factors Roundtable Part 1: The Regulatory Imperative. MDDI - Medical Device & Diagnostic Industry Magazine. Abruf unter: <http://www.devicelink.com/mddi/archive/01/01/006.html> (18.10.2007)
- 417 Willumeit H., Gediga G., Hamborg K. (1996): IsoMetrics(L). Ein Verfahren zur formativen Evaluation von Software nach ISO 9241/10. In: Ergonomie und Informatik. 27. 1996. S. 5-12
- 418 Wilson R. M., Runciman W. B., Gibberd R. W., Harrison B.T., Newby L., Hamilton J.D. (1995): The Quality in Australian Health Care Study. In: Med J.Aust 163 (9) 1995, S. 458-471.

- 
- 419 Wittekind Ch., Meyer H.-J., Bootz F.(Hrsg.) (2002) : UICC: TNM Klassifikation maligner Tumoren. 6. Auflage. Berlin, Heidelberg, New York. Springer
- 420 Woodson W. E. (1981): Human Factor Design Handbook. New York. Mc Graw Hill Book Company.
- 421 Wottawa H, Thierau H. (1998): Lehrbuch Evaluation. 2. Auflage. Bern. Verlag Hans Huber.
- 422 Wottawa H. (2001): Evaluation. In: Weidemann B., Krapp A.: Pädagogische Psychologie 4., vollst. überarb. Aufl. Weinheim. Beltz PVU
- 423 Young M., Stanton N. (2005): Applying Interviews to Usability Assessment. In: Stanton N., Hedge A., Brookhuis K., Salas E., Hendrik H.: Handbook of Human Factors and Ergonomics Methods. Boca Ration FL. CRC Press. S. 29-1 - 29-6
- 424 Zangemeister C. (1976): Nutzwertanalyse in der Systemtechnik. Eine Methodik zur multi-dimensionalen Bewertung und Auswahl von Projektalternativen. Diss. Techn. Univ. Berlin 1970, 4. Aufl., München: Wittemann
- 425 Zentralstelle der Länder für Gesundheitsschutz bei Arzneimitteln und Medizinprodukten (2007): Normen und technische Vorschriften 3.5 A 2 Ergonomie
- 426 Zhang J., Johnson T.R., Patel V.L., Paige D.L., Kubose T. (2003): Using usability heuristics to evaluate patient safety of medical devices Journal of Biomedical Informatics 36 (2003) S.23–30
- 427 Zhang J., Norman D. A. (1994): Representations in distributed cognitive tasks. Cognitive Science. 18(1), 87-12
- 428 Zhang J., Patel V. L., Johnson T. R., Chung P. H., Turley J. P. (2005). Evaluating and predicting patient safety for medical devices with integral information technology. In Henriksen K., Battles J. B., Marks E., Lewin D. I. (Eds.): Advances in Patient Safety: From Research to Implementation. Rockville, MD: Agency for Healthcare Research and Quality. S. 323-336
- 429 Zhang J., Patel V. L., Johnson T. R., Shortliffe E. H. (2002): Toward An Action Based Taxonomy of Human Errors in Medicine. In: Annual Conference of Cognitive Science Society. Mahwah. Lawrence Erlbaum Associates. S. 970-975.
- 430 Zhang J., Patel V.L., Johnson K.A., Malin J., Smith J.W.: Designing human-centered distributed information systems. IEEE Intell Syst 2002;17(5):42–7.
- 431 Zimmer A. (2001): Wie intelligent darf/muss ein Auto sein? Anmerkungen aus ingenieurspsychologischer Sicht. In: Jürgensohn T., Timpe K.-P. (Hrsg.): Krafffahrzeugführung. Berlin: Springer. S. 263-28
- 432 Zimolong A. (2004): Verfahren zur Zuverlässigkeitsbeurteilung der Mensch-Maschine-Interaktion bei computergestützten Chirurgesystemen. Rau G. (Hrsg.): Berichte aus der Biomedizinischen Technik 17. Aachen. Shaker
- 433 Zipper S.G. (2006): Medical Risk Management. Med Klin 2006;101, (Nr. 10). S. 796–803
- 434 Zühlke.(2005): Der intelligente Versager. Das Mensch-Technik-Dilemma. Darmstadt. Primus Verlag

## Anhang

## Anhang A: Zusammenstellung verschiedener Fragebögen (vgl. Kap. 2.4.1)

Fragebogen	Beschreibung
ASQ (After Scenario Questionnaire)	<p>Von Lewis entwickelter Fragebogen, welcher in einem szenarienbasierten Gebrauchstauglichkeitstest nach jedem Szenario mit drei Items die Nutzerzufriedenheit ermittelt. Die Items richten sich an die drei für die Nutzerzufriedenheit wichtigen Komponenten: Einfachheit der Aufgabenerfüllung, Bearbeitungszeit und angemessene Supportinformation (Hilfen, Hinweise, Dokumentation). Die Urteile erfolgen mit einer 7-stufigen bipolaren Likert-Skala mit verbalen Marken und einer Auswahlmöglichkeit „Nicht anwendbar“ [242, 243].</p> <div data-bbox="576 712 1353 1021" style="border: 1px solid black; padding: 5px;"> <p><b>Ich bin zufrieden damit, wie einfach ich die zuletzt gestellte Aufgabe lösen konnte.</b></p> <p>Stimme voll zu <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> Stimme gar nicht zu <input type="checkbox"/> Nicht anwendbar</p> <hr/> <p><b>Ich bin zufrieden mit der benötigten Bearbeitungszeit für diese Aufgabe.</b></p> <p>Stimme voll zu <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> Stimme gar nicht zu <input type="checkbox"/> Nicht anwendbar</p> <hr/> <p><b>Ich bin zufrieden mit der angebotenen Hilfeinformation bei dieser Aufgabe.</b></p> <p>Stimme voll zu <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> Stimme gar nicht zu <input type="checkbox"/> Nicht anwendbar</p> </div> <p>Sauro und Kindlung beschreiben einen ähnlichen [351] Fragebogen, welchen Sie „Post-task questionnaire“ nennen. Dieser Fragebogen beinhaltet neben verschiedenen verbalen Marken und einer lediglich fünfstufigen Likert-Skala als größten Unterschied, dass das Item bzgl. der Hilfeinformation durch ein Item ersetzt wurde, welches die Zufriedenheit mit der Verwendung der Applikation bewertet.</p> <p><b>Fazit:</b> Der ASQ ist ein kurzer Fragebogen, der ein direktes Feedback der Nutzerzufriedenheit zu einem Aufgabenszenario gibt. Er ist einfach und schnell zu verwenden, standardisiert und in der Literatur etabliert.</p>
PSSUQ (Post-Study System Usability Questionnaire)	<p>Von Lewis entwickelter Fragebogen zur Ermittlung der Nutzerzufriedenheit bezüglich der Gerätegebrauchstauglichkeit nach einem Nutzertest. Der Fragebogen besitzt 19 Items, welche von einer Gruppe von Gebrauchstauglichkeitsexperten entwickelt wurden und sich auf die Systemeigenschaften: Einfach zu bedienen, einfach zu erlernen, Einfachheit, Effektivität, Informationen und die Benutzerschnittstelle beziehen. Die Urteile erfolgen mit einer 7-stufigen bipolaren Likert Skala mit verbalen Marken und einem „Nicht anwendbar“-Punkt außerhalb der Skala [243, 239, 244]. Die Ergebnisse des PSSUQ können in vier Bereiche unterteilt werden: overall satisfaction score (OVERALL): Item 1-19, system usefulness (SYSUSE) Item 1-9, information quality (INFOQUAL) Item 9-15 und interface quality (INTERQUAL): Item 16-18.</p> <p><b>Fazit:</b> Der PSSUQ ist als Fragebogen nach dem Test geeignet, da er passende Themenschwerpunkte und übertragbare Frageformulierungen bietet. Er ist ein kompakter, standardisierter und etablierter Fragebogen</p>

SUMI (Software Usability Measurement Inventory)	<p>Von Kirakowski und Corbett entwickelter Fragebogen, um Softwarequalität aus dem Blickwinkel des Endnutzers zu messen [216]. Der SUMI besteht aus 50 Aussagen, welche der Nutzer mit „ich stimme zu“, „ich stimme nicht zu“ oder „ich weiß nicht“ beantworten muss. In der DIN EN 9241-11 [113] wird der SUMI als eine Methode vorgeschlagen, um Nutzerzufriedenheit zu messen.</p> <p><b>Fazit:</b> Der SUMI bewertet das Gesamtsystem. Durch den fehlenden Aufgabenbezug des Fragebogens ist im Anschluss an das Ausfüllen ein Interview nötig, um detaillierte Angaben zu Problemen zu erhalten. Studien zeigen beim SUMI eine befriedigende Reliabilität. Die Validierung wurde in verschiedenen Untersuchungen vorgenommen, in denen dieser Fragebogen zeigen konnte, dass er zwischen Software unterschiedlicher ergonomischer Qualität unterscheiden kann [125].</p>
SUS (System Usability Scale)	<p>Von Brooke entwickelter, aus 10 Items bestehender Fragebogen, welcher einen allgemeinen Überblick der Nutzerzufriedenheit bezüglich der Gebrauchstauglichkeit eines Systems gibt. Die Urteile erfolgen mit einer 5-stufigen, bipolaren Likert-Skala mit verbalen Marken an den Polen. Die Antworten ergeben mit Hilfe einer Berechnungsformel einen SUS-Wert zwischen 0 und 100. „100“ bedeutet hierbei eine hohe Nutzerzufriedenheit [40].</p> <p><b>Fazit:</b> SUS ist ein einfacher und schneller Fragebogen, der jedoch keine Informationen zu genauen Problemen liefert. Hier wäre ein weiteres Interview nötig.</p>
QUIS (Questionnaire for User Interaction Satisfaction)	<p>Von einer interdisziplinären Expertengruppe entwickelter Fragebogen zur Messung der Nutzerzufriedenheit bei der Systeminteraktion. Der Fragebogen umfasst in seiner Kurzform fünf Kategorien mit 46 Items. Nutzer beurteilen auf einer 9-stufigen bipolaren Skala Terminologie, Systeminformationen (Systemstatus, Anweisungen, Fehlermeldungen etc.), Erlernbarkeit des verwendeten Systems und Systemfähigkeiten (Schnelligkeit, Zuverlässigkeit usw.). Weiterhin beinhaltet der Fragebogen noch einen demografischen Teil [169, 368].</p> <p><b>Fazit:</b> Die Reliabilität des Gesamtergebnisses wurde von Chin et al. überprüft und für gut befunden [61], jedoch wurden keine Untersuchungen zu den einzelnen Kategorien durchgeführt.</p>
ISOMetrics	<p>Der von Willumheit, Gediga und Hamborg entwickelte Fragebogen ISOMetrics ist eine Operationalisierung des Standards DIN EN 9241-10. Er dient der benutzerorientierten Evaluation der Grundsätze der Dialoggestaltung von grafischen Benutzungsschnittstellen entsprechend dieser Norm. Zu jedem dieser Kriterien existieren verschiedene Aussagen, die in einer fünfstufigen bipolaren Ratingskala mit verbalen Marken die Zustimmung oder Ablehnung der Anwender verlangen. Der Fragebogen mit 75 Items ist in einer kurzen und in einer langen Version verfügbar. Beide Versionen verwenden dieselben Items. Bei der langen Version wird die zu bewertende Aussage noch durch eine Bewertung der Wichtigkeit des Items für den</p>

Gesamteindruck und einer Aufforderung zur Beispielnennung ergänzt [417, 166].

		stimmt nicht	stimmt wenig	stimmt mittelmäßig	stimmt ziemlich	stimmt sehr	Keine Angabe
A.6	Die Software ermöglicht es mir, Daten so einzugeben, wie es von der Aufgabenstellung gefordert wird.	1	2	3	4	5	

**Fazit:** Der ISOMetrics-Fragebogen ist auf die internationale Norm DIN EN 9241 Teil 10 ausgerichtet. Der Fragebogen ist von Gediga et al. auf seine Zuverlässigkeit und Validität geprüft worden [148]. Es werden konkrete Informationen über Fehler und Probleme geliefert. Hamborg zeigt, dass diese Methode im Vergleich zu anderen effektiv ist, um Gebrauchstauglichkeitsprobleme zu evaluieren. Dieser Fragebogen ist jedoch auf Softwareprodukte ausgerichtet und müsste für eine Gebrauchstauglichkeits-evaluation für medizinische Geräte adaptiert werden. Weiterhin ist dieser Fragebogen nicht auf einzelne Aufgaben ausgerichtet.

ISONORM 9241/10

Von Prümper und Anft entwickelter Fragebogen, welcher die Konformität mit den Dialogprinzipien entsprechend der DIN EN 9241-10 prüft. Zu jedem der Dialogprinzipien werden fünf Aussagen mit jeweils einem positiven und einem negativen Pol getroffen, welche die Nutzer auf einer siebenstufigen Likert-Skala von vollkommener Ablehnung bis hin zur vollkommenen Zustimmung bewerten können. Insgesamt umfasst der Fragebogen somit 35 Items [317, 318].

<b>Aufgabenangemessenheit</b>								
Unterstützt die Software die Erledigung Ihrer Arbeitsaufgaben, ohne Sie als Benutzer unnötig zu belasten?								
Die Software...	---	--	-	-/+	+	++	+++	
ist kompliziert zu bedienen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	ist unkompliziert zu bedienen.
bietet nicht alle Funktionen, um die anfallenden Aufgaben um effizient zu bewältigen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	bietet alle Funktionen, die anfallenden Aufgaben effizient zu bewältigen.
(...)								

**Fazit:** Dieser Fragebogen ist auf die internationale Norm DIN EN 9241 Teil 10 ausgerichtet. Es werden keine konkreten Informationen über Fehler und Probleme geliefert, da er die einzelnen Dialogprinzipien lediglich bewertet und nicht die einzelnen Aufgaben fokussiert.



ErgoNorm-  
Benutzer-  
fragebogen  
zu "Arbeit u.  
Software"

Mit dem von Dzida et al. entwickelten Fragebogen der „Bundesanstalt für Arbeitsschutz und Arbeitsmedizin“ kann eine Software subjektiv bewertet werden. Er basiert auf der Grundlage der DIN EN 9241 Teil 10 und 11 und wird zur Konformitätsprüfung herangezogen. Dem Nutzer werden zu den einzelnen Dialogprinzipien Fragen gestellt, welche er mit Ja oder Nein beantworten kann. Bei einer negativen Aussage soll er zusätzlich Erläuterungen abgeben. Insgesamt besitzt der Fragebogen 18 Items [128].

**Müssen Sie oft Kollegen oder ein Handbuch konsultieren, um weiterarbeiten zu können?**

 ja nein Frage trifft nicht zu

**wenn "ja":**

Nennen Sie bitte Situationen, in denen Sie auf die Hilfe von Kollegen oder eines Handbuchs angewiesen waren.

---

---

---

 Ich empfinde dies als sehr störend

Fazit: Der ErgoNorm-Benutzerfragebogen ist auf die internationale Norm DIN EN 9241 Teile 10 und 11 ausgerichtet. Durch die Beispielnennung können konkrete Informationen über Fehler und Probleme evaluiert werden. Da dieser Fragebogen jedoch wenig detailliert und auf keine konkreten Aufgaben ausgerichtet ist, wird er eher oberflächlich ein Produkt bewerten können.

## Anhang B Zusammenfassung der in dieser Arbeit betrachteten Gebrauchstauglichkeitsevaluationsmethoden (vgl. Kap. 2.4.1)

	Inspektionsmethoden						
	Cognitive Walkthrough	Pluralistic Walkthrough	Heuristische Evaluation	Funktionsanalyse	Konsistenzprüfung	Checklisten	Fokusgruppe
Ziel	Experten simulieren hierbei einen Nutzer und erkunden so das Gerät und dessen Bedienung.	Interdisziplinäres Team (inkl. Nutzer) versucht, Gebrauchstauglichkeitsprobleme anhand von Szenarien zu entdecken.	Überprüfung einer Benutzerschnittstelle anhand von Heuristiken.	Typische Funktionalitäten (engl. features) auf mögliche Probleme hin untersuchen	Bei komplexen Geräten, Geräteverbänden oder ganzen Systemen verwendet, um Konsistenz und Einheitlichkeit innerhalb des Produktes und des Systems zu überprüfen	Erfassung einer Benutzerschnittstelle anhand von Prüfpunkten einer Checkliste	Identifizierung erster Anforderungen [349] Eine Gruppe von Teilnehmern bewertet ein Produktkonzept oder ein bestehendes Produkt. Sie kann aber auch zur Nutzungskontextanalyse verwendet werden.
Zeitpunkt	Formativ [350]	Formativ oder Summativ	Formativ oder Summativ [350]	Formativ [190]	Formativ oder Summativ	Formativ oder Summativ	Aufgabenanalyse [284] Formativ oder Summativ
Benötigte Zeit	Mittel [190, 350]	Mittel	Gering [190, 350]	Mittel [190]	Gering bis Mittel (je nach System)	Gering (Vorarbeit notwendig)	Gering
Benutzerbeteiligung	Nein [190]	1+	Nein [190, 284]	Nein [190]	Nein	Nein	6-9 pro Gruppe [284]
Benötigte Evaluatoren	3+ [190]	3+ (interdisziplinär)	3+ [190]	1-2 [190]	1-2	1-2	1 Leiter
Benötigte Technik	Gering [190]	Gering	Gering [190]	Gering [190]	Gering	Gering	Gering
Benötigte Erfahrung	Hoch [190], Mittel [350]	Hoch	Medium [190, 350]	Hoch [190]	Medium	Medium	Mittel

	Empirische Methoden				
	Teilnehmende Beobachtung	Feldbeobachtung	Coaching Methode	Nutzertest	Benutzerbefragung
Ziel	Beobachter ist Teil des Geschehens und beobachtet den Nutzer in seinem Nutzungskontext.	Beobachtung in der realen Nutzungsumgebung	1. Nutzer stellt Fragen an den Coach oder 2. Benutzer erklärt Experten die Benutzerschnittstelle	Nutzertest im Labor	u.a. Erfassen der subjektiven Zufriedenheit und Festlegen von Ursachen für Unzufriedenheit [349]
Zeitpunkt	Formativ in der Aufgabenanalyse [284] Summativ, erst wenn Produkt verwendet wird	Summativ [190], Formativ: Ermittlung des Nutzungskontext, Summativ: erst wenn fertiges Produkt vorliegt	Formativ und summativ	Beides [190, 350], Iteratives design/formativ [284]	Beides [190, 350], Aufgabenanalyse u. Nachfolgestudien [284]
Benötigte Zeit	Mittel	Mittel [190]	Mittel	Hoch [190, 350]	Gering bis Mittel [190], Gering [350] (Vorarbeit notwendig)
Benutzerbeteiligung	Ja (hängt von der Anwendugn ab)	10+ [190]	1+	6+ [190], 3-5 [284]	30+ [190, 284]
Benötigte Evaluatoren	1-2, 3+ [284]	1+ [190]	1	1-3 [190]	1+ [190]
Benötigte Technik	Gering	Mittel [190]	Gering	Hoch [190]	Gering [190]
Benötigte Erfahrung	Mittel	Hoch [190]	Mittel bis Hoch	Medium [190], Gering [350]	Gering [190, 350]

## Anhang C - Zusammenfassung verschiedener Studien zu Gebrauchstauglichkeitsevaluationsmethoden (vgl. Kap. 2.4.2)

Studie	Inhalt	Methode	Ergebnis	Bewertung
Desurvire (1994) [72]	HE u. CWT vs. NT	Vergleich der Anzahl der gefundenen Fehler	Mit Gebrauchstauglichkeits-Inspektionsmethoden können Gebrauchstauglichkeitsprobleme aufgedeckt werden. Sie sind besonders wirksam, wenn Sie von Experten verwendet werden, können jedoch NTs nicht ersetzen	1. NT 2. HE mit Experten 3. CWT mit Experten
Doubleday et al. (1997) [120]	HE vs. NT	Vergleich der Fehlerart und -anzahl	HE und NT können unterschiedliche Probleme aufdecken. NTs identifizieren die Symptome der Probleme, HE checkt die Ursache und kann mögl. Lösungen aufdecken.	NT = HE
Jeffries et al. (1991) [205]	HE, NT, CWT, GL	Vergleich der Fehlerart und -anzahl	HE liefert die meisten Ergebnisse. Guidelines und CWT sind eher Techniken für Designer, um schnell einige Probleme zu identifizieren, wenn keine Usability-Experten vorhanden sind. NT und HE finden mehr und vor allem schwerwiegende Probleme.	1. HE 2. NT 3. GL 4. CWT
Desurvire et al. (1991) [74]	HE (Experten, Laien) Vs. NT	R <sup>2</sup>	HE ist geeignet, um Aufgaben zu identifizieren, in denen spezifische Probleme auftreten.	1. NT 2. HE mit Experten 3. HE mit Laien
Karat et al. (1992) [210]	NT, WT (2er-Team, Einzelsitzung)	Vergleich der Fehlerart und -anzahl	NT und Team Walkthrough haben Vorteile gegenüber dem Einzel-Walkthrough. Mit NT konnten 4mal mehr signifikante Probleme gefunden werden als beim Teamwalkthrough, und die Teams finden 33-50% mehr Informationen als der Einzel-Walkthrough	1. NT 2. WT im Team 3. WT
Nielsen (1990) [282]	HE mit Laien, Experten und Doppel-experten	Vergleich der Fehlerart und -anzahl	Ergänzt vorherige Studien über die Leistung einer Heuristischen Evaluation und Empfehlungen von Gruppengrößen von 4-6 Personen um die Anpassung der Gruppengröße entsprechend der Erfahrung der Evaluatoren. Eine kleine Gruppe kann bei Doppelexperten angewendet werden, bei Laien hingegen muss die Gruppengröße erhöht sein.	1. HE mit Doppel-experten 2. HE mit Usability Experten 3. HE mit Laien
Nielsen (1995) [285]	NT, HE, FI, HES, CI, SI, PW	Befragung von Methoden-anwendern	NT und HE werden am meisten angewandt. Jedoch liefern alle Methoden Ergebnisse.	1. NT 2. HE 3. HES 4. FI 5. CI 6. CWT 7. PW 8. SI
Dutt, Johnson und Johnson (1994) [126]	HE, CWT	Vergleich der Anzahl der gefundenen Fehler	Methoden ergänzen sich gegenseitig. Methoden sollten jedoch in unterschiedlichen Phasen des Designprozesses angewendet werden.	HE = CWT
Desurvire et al. (1992) [73]	NT, HE, CWT (Experten, Softwareentwickler, Laien)	Anzahl der gefundenen Fehler	NT findet die meisten Probleme. HE ist CWT überlegen. Bei den Inspektionsmethoden ist die mit Experten die am effizienteste.	1. NT 2. HE mit Experten 3. CWT mit Experten 4. HE mit Entwicklern, CWT mit Entwicklern 5. HE mit Nicht-Experten, CWT mit Nicht-Experten
Fu et al. [142]	NT, HE	Anzahl der gefundenen Fehler	Viele der in der HE gefundenen Probleme stellen sich als nicht relevant heraus.	1. HE 2. NT
HE = Heuristische Evaluation GL = Guidelines (Richtlinien) FI= Feature Inspection SI= Standards Inspection UAN= User Action Notation CWT = Cognitive Walkthrough WT = Walkthrough HES = Heuristic Estimation PW= Pluralistic Walkthrough NT = Nutzertest Spec.= Reading Specification CI= Consistency Inspection CA= Claim Analysis				

## Anhang D – Thesenverifikation durch die Evaluationsstudien

		Studie: HF (vgl. Kap. 4.2.1.)	Studie: MIC (vgl. Kap. 4.2.2.)	Studie: Spritzenpumpe (vgl. Kap. 4.2.3)	Studie: Herzkatheter- pumpe (vgl. Kap. 5.5)
	<b>Anforderungsanalyse:</b>				
T.1	Der medizinische Nutzungskontext ist durch Benutzer, Arbeitsaufgabe, Arbeitsmittel und Arbeitsumgebung bestimmt.	Ja	Ja	Ja	Ja
T.2	Für die Überprüfung der Gebrauchstauglichkeit medizinischer Geräte ist die Erfassung der beeinflussenden Faktoren des Nutzungskontextes unerlässlich.	Ja, sehr wichtig	Ja	Ja	Ja
	<b>Nutzerprofil:</b>				
T.3	Methoden zur Entwicklung des Nutzerprofils sind Literaturrecherche, Beobachtung, Nutzerinterview, Fragebogenbefragung, Hersteller- und Vertrieberinformationen oder Experteninterviews.	Ja, Hersteller und Anwender weichen ab.	Die Methodenwahl ist auf den speziellen Fall abzustimmen.	Ja	Ja
T.4	Eine Ermittlung und Unterteilung der Nutzergruppen ist notwendig.	Ja	Ja, auch zwischen Novizen und Experten	Ja. Expertennutzer haben nicht unbedingt weniger Probleme.	Ja. Expertennutzer haben nicht unbedingt weniger Probleme.
T.5	Unterschiedliche Nutzer medizinischer Geräte benutzen diese anders und unterschiedliche Ergebnisse werden evaluiert.	Ja, in der Aufgabenanalyse erkennbar	Ja. Novizen und Laien zeigen unterschiedliche Ergebnisse. Die Unterschiede zwischen den Geräten sind jedoch vergleichbar.	Hier treten vergleichbare Probleme auf. Probleme treten auch bei der Laiengruppe auf.	Vergleichbare Probleme. Probleme treten auch bei der Laiengruppe auf
	<b>Aufgabenprofil:</b>				
T.6	Als Methoden für ein Aufgabenprofil kommen Befragungen (Interviews, Fragebögen) von Nutzern und Herstellern, Beobachtungen des Anwenders, Brainstorming-Techniken, Fokusgruppen und/oder Kontextbegehungen in Betracht.	Befragung von Nutzern und Herstellern, Brainstorming,	Befragung von Experten	Ja	Ja
T.7	Mit der Aufgabenanalyse können die mit dem System durchgeführten Ziele, Aufgaben und Aktionen der Nutzer ermittelt und kategorisiert werden. Sie liefert somit elementare Informationen für die Gebrauchstauglichkeitsprüfung.	Ja	Ja, HTA ist sehr übersichtlich	Ja	Ja
T.8	Die Aufgaben sollten für die Evaluation in Primär- und Sekundäraufgaben unterteilt werden.	NB	Ja	Ja	Ja
	<b>Umgebungsanalyse:</b>				
T.9	Die Analyse der Arbeitsumgebung und der Arbeitsmittel liefert elementare Informationen für den Aufbau der Gebrauchstauglichkeitsprüfung.	Ja	Ja. In gut untersuchten Bereichen können eine Literaturrecherche und eine Umgebungsbegehung ausreichen.	Ja	Ja
T.10	Die Umgebungsanalyse liefert Informationen darüber, welche Faktoren als leistungsbeeinflussend bei der Gebrauchstauglichkeitsevaluation berücksichtigt werden müssen.	Ja	Ja	Ja	Ja
	<b>Prüfkriterien – Wie misst man Usability? Messung der Gebrauchstauglichkeit</b>				
T.11	Aus Nutzeranalyse, Anforderungen an das Gerät und einer Expertenevaluati- on können Testkriterien abgeleitet werden.	Ja	Ja	Ja	Ja
T.12	Usability Test und Heuristische Evaluation liefern die meisten Probleme.	Ja	Ja	Ja. Hier wurden keine	Ja

				Heuristiken verwendet, scheinbar wurden weniger Probleme gefunden.	
T.13	Die Heuristische Evaluation sollte für medizinische Geräte angepasst werden.	Ja	Ja	Ja	Ja
T.14	Die Heuristische Evaluation erfasst potentielle Probleme, welche im Gebrauchstauglichkeitstest überprüft werden können. Hierzu sollten entsprechende Aufgaben gewählt werden.	Ja	Ja	NB	Ja
T.15	Zur Messung der Gebrauchstauglichkeit in der Medizin werden Leistungsmaße und subjektive Maße benötigt.	Ja. Aber zusätzlich muss auch die Sicherheit betrachtet werden	Ja. Bei mechanischen Tätigkeiten und potentiellen körperlichen Belastungen sollten auch Beanspruchungsmessungen durchgeführt werden.	Ja	Ja
T.16	Leistungsmaße können in Zusammenarbeit mit Expertennutzern aufgestellt werden.	Ja	Ja	Ja. Der Experte verwendet das Gerät. Die Zeit, welche bei einer Aufgabenerfüllung ohne Problem ermittelt wird, kann als Referenz dienen.	Ja
T.17	Für die Erhebung subjektiver Zufriedenheitsdaten können Fragebögen eingesetzt werden. Mit einem aufgabenspezifischen Fragebogen (z.B. ASQ), speziellen Fragen zu den einzelnen Hauptaufgaben mit dem Gerät und der Erfassung des Gesamteindrucks ist eine Bewertung der Gebrauchstauglichkeit möglich.	Ja	Ja. Aber auch Fragen speziell zu den Arbeitsaufgaben und einzelnen Aktionen sind zu stellen und eine Gesamtbewertung ist abzugeben.	Ja	Ja
T.18	Um die Gebrauchstauglichkeit in der Medizin zu bewerten, reichen die Maße Effektivität, Effizienz und Zufriedenheit nicht aus, es muss zusätzlich die Sicherheit und das Risiko betrachtet werden.	Ja	Ja	Ja	Ja
<b>Evaluatoren</b>					
T.19	Die Standardprozedur hilft, den Evaluatoreffekt zu vermeiden, da sie strukturiert den Nutzungskontext und die Ziele und Aufgaben analysiert, verschiedene Evaluatoren mit einbezieht und einen standardisierten Evaluationsleitfaden vorgibt.	NB	NB	NB	Ja
T.20	Die Heuristische Evaluation sollte mit Doppelsexperten und Fachexperten durchgeführt werden.	Ja	Ja	NB	Ja
T.21	Die Bewertung sollte mit mehreren Doppel- und Fachexperten durchgeführt werden.	Ja	NB	Ja	Ja
<b>Stichprobe</b>					
T.22	Die Stichprobe sollte ein ausgewogenes Verhältnis von Laien zu Fachnutzern aufweisen.	Ja	Ja, auf jeden Fall. Die Studie zeigt, dass es zu unterschiedlichen Ergebnissen kommt.	Hier hätte auch keine Unterscheidung die Probleme evaluiert.	Ja
T.23	Für die Stichprobe sind Probanden zu nehmen, welche dem Nutzerprofil entsprechen.	Ja, aber auch Novizen mit Vorwissen.	Ja, nur eine ausgewogene Stichprobe führt zu einer kompletten Betrachtung.	Ja	Ja
T.24	Die Anzahl der Nutzer muss bei der Gebrauchstauglichkeitsprüfung in der Medizin höher als 6 sein. Die Problemfindungsrate sollte bei mindestens 90%	Ja. Bei 6 wären bei ERBE nur ca. 71%	Hier liegt bei 6 Nutzern die Problemfindungsrate bei 80,5%.	Hier liegt bei 6 Nutzern die Problemfindungsrate bei	Hier liegt bei 6 Nutzern die Problemfindungsrate

	liegen.	der Fehler entdeckt worden.		91%.	bei 66,3%.
T.25	Mit einer Stichprobengröße von 9 Nutzern werden mehr als 80% der Probleme gefunden. Hiermit ist meist eine Bewertung möglich.	Ja, aber hier auch schon bei ca. 8 Nutzern. Jedoch gilt dies nur für die gewählten Aufgaben	Ja, aber hier schon bei 6 Nutzern. Bei 9 liegt die Problemfindungsrate bei 91,4%.	Ja, aber auch schon bei 4 Nutzern. Bei 6 Nutzern liegt die Problemfindungsrate bei 91,9% und bei 9 bei 97,7%	Ja, aber nur knapp mit 80%, daher eher 10 Nutzer einsetzen.
<b>Aufgaben und Szenarien</b>					
T.26	Anwendungsszenarien helfen bei der Definition des Nutzungskontextes und der Erstellung der Testaufgaben.	Ja	Ja	Ja	Ja
T.27	Aufgabenszenarien vermitteln Nutzern in natürlicher Weise Aufgabenziel und Informationen für die Aufgabe.	Ja, jedoch reicht dies alleine nicht aus, auch die Testumgebung muss stimmen	Ja	Ja	Besser ist natürlicher Dialog.
T.28	Aus den Anwendungsszenarien, den Kontextszenarien und den Informationen der Anforderungsanalyse können die für die Evaluation notwendigen repräsentativen Testaufgaben abgeleitet werden.	Ja	Ja	Ja	Ja
T.29	Testaufgaben sollen hauptsächlich Primäraufgaben enthalten. Es sollen aber auch Sekundäraufgaben und Aufgaben in Notsituationen gestellt werden sowie Aufgaben, welche in der Anforderungsanalyse oder in einem Expertenreview als potentiell problematisch oder risikoträchtig identifiziert wurden. Zusätzlich soll auch, entsprechend den Forderungen der DIN EN 60601-1-6, ein vernünftigerweise vorhersehbarer Missbrauch betrachtet werden.	Ja	Ja	Ja	Ja
<b>Testaufbau:</b>					
T.30	Gebrauchstauglichkeitsuntersuchungen in der Medizin müssen in einer Laborumgebung durchgeführt werden, in welcher der Nutzungskontext simuliert werden kann.	Ja	Ja	Ja	Ja
<b>Report:</b>					
T.31	Der für Softwareprodukte standardisierte Gebrauchtauglichkeitsbericht kann in einer angepassten Form für Medizinprodukte verwendet werden.	-	Ja	Ja	Ja
<b>Testgüte</b>					
T.32	Die Standardprozedur erfüllt die klassischen Testgütekriterien.	Ja	NB	NB	Ja
T.33	Objektivität wird durch die Standardprozedur erreicht. Die Korrelation zwischen zwei Testanwendern sollte bei einem Wert größer 0,8 liegen.	NB	NB	NB	Ja, nach einer Bereinigung der Ergebnisse
T.34	Eine hohe Reliabilität kann durch die Standardprozedur erreicht werden. Ein Test nach wiederholter Anwendung führt zu den annähernd gleichen Ergebnissen ( $r=0,75$ ).	Ja	Ja, hier durch standardisierte Vorgehensweise und Bericht	NB	NB
T.35	Zumindest bei den schwerwiegendsten Problemen sollte in einem wiederholten Test die gleiche Benotung herauskommen.	Ja	NB	NB	Ja
T.36	Standardisierte Methoden und eine repräsentative Stichprobe sichern die Validität.	Ja	Ja	Ja	Ja

NB = Nicht betrachtet

---

## Anhang E – Testaufgaben für ein HF Gerät (vgl. Kap. 4.2.1 )

**Aufgabe 1:** Im OP ist ein HF-Gerät ausgefallen. Um die OP durchführen zu können, wurde im Nachbarsaal ein HF-Gerät abgebaut. Sie sollen nun das HF-Gerät in Ihrem OP-Saal wieder aufbauen und in Betrieb nehmen.

**Aufgabe 2:** Der operierende Chirurg möchte für eine Gefäßversiegelung die bipolare Pinzette verwenden. Weisen Sie dem Fußschalter die Aktivierung der bipolaren Pinzette zu.

**Aufgabe 3:** Der operierende Chirurg möchte mit dem HF-Gerät arbeiten, hat Ihnen aber keine genauen Anweisungen gegeben, welche Einstellungen bzw. welches Programm er verwenden möchte. Wählen Sie deshalb an dem HF-Gerät das Programm: „Standard“ aus.

**Aufgabe 4:** Der operierende Chirurg gibt Ihnen genaue Anweisung welche Werte er am Gerät eingestellt haben will, um damit einen Schnitt und eine Koagulation durchzuführen. Stellen Sie an dem HF-Gerät die gewünschten Werte und Einstellungen ein.

**Aufgabe 5:** Der leitende Operateur hat am Patienten einen Schnitt im Gewebe vorgenommen. Er gibt Ihnen das monopolare Instrument und weist Sie an, diesen Schnitt an einer anderen Stelle zu wiederholen.

**Aufgabe 6:** Der leitende Operateur hat am Patienten eine Koagulation am Gewebe vorgenommen. Er gibt Ihnen das HF-Instrument und weist Sie an diese Koagulation an einer anderen Stelle zu wiederholen.

**Aufgabe 7:** Der leitende Operateur hat beim Patienten eine Gefäßversiegelung unter Verwendung der bipolaren Pinzette vorgenommen. Er gibt Ihnen die bipolare Pinzette und weist Sie an, mit dieser noch ein weiteres Gefäß zu versiegeln.

**Aufgabe 8:** Der operierende Chirurg gibt Ihnen die Anweisung, die Auto-Start und Auto-Stopp-Funktion des HF-Gerätes zu aktivieren und diese der bipolaren Pinzette zuzuordnen.

**Aufgabe 9:** Der operierende Chirurg gibt Ihnen Anweisungen welche Einstellungen Sie vornehmen müssen, damit er seine beabsichtigte Anwendung ausführen kann. Am Gerät wurde ein technischer Fehler erzeugt, der zuvor behoben werden müsste.

**Aufgabe 10:** Der operierende Chirurg weist Sie an, das Programm (z.B. Chirurgie) auszuwählen, um mit diesen Einstellungen einen Schnitt auszuführen.

**Aufgabe 11:** Der operierende Chirurg weist Sie an, die Werte, wie er sie Ihnen gegeben hat, als Programm für ihn unter dem Namen „Test“ abzuspeichern, damit bei der nächsten OP wieder darauf zurückgegriffen werden kann.

**Aufgabe 12:** Die OP ist erfolgreich beendet worden. Sie werden vom Chirurgen angewiesen, das Gerät abzuschalten und abzubauen, damit die Instrumente gereinigt und sterilisiert werden können.

## Anhang F: Problemfindungsrate und Berechnung des angepassten p-Werts (Adj p) für die HF-Geräte (vgl. Kap. 4.2.1)

### HF-Gerät 1

Problem	Test (Teilnehmernummer)										Frequenz (in %)
	1(1/2)	2(3)	3(4)	4(5/6)	5(7/8)	6(9)	7(10/11)	8(12/13)	9(14)	10(15/16)	
1											20
2											30
3											40
4											50
5											30
6											70
7											70
8											100
9											20
10											10
11											10
12											10
13											20
14											20

n = 10

$$\text{est } p = (0,2+0,3+0,4+0,5+0,3+0,7+0,7+1+0,2 + 0,1+0,1+0,1+0,2+0,2)/14 = 0,357142857$$

GTadj = 3/14 = 0,2143

est p = .3571

Adj p = 0,5\*[(0,3571 - (1/10))\*(1-(1/10))]+0,5\*(0,3571/(1+0,2143))] = 0,2627

### HF-Gerät 2

Problem	Test (Teilnehmernummer)										Frequenz (in %)
	1(1/2)	2(3)	3(4)	4(5/6)	5(7/8)	6(9)	7(10/11)	8(12/13)	9(14)	10(15/16)	
1											60
2											30
3											90
4											100
5											80
6											50
7											100
8											30
9											80
10											40
11											100

n=10

$$\text{est } p = (0,6+0,3+0,9+1+0,8+0,5+1+0,3+0,8+0,4 + 1)/11 = 0,6909$$

GTadj = 0/11 = 0

est p = .6909

Adj p = 0,5\*[(0,6909-(1/10))\*(1-(1/10))]+0,5\*(0,6909/(1+0))] = 0,6113

### HF-Gerät 2

Problem	Test (Teilnehmernummer)										Frequenz (in %)
	1(1/2)	2(3)	3(4)	4(5/6)	5(7/8)	6(9)	7(10/11)	8(12/13)	9(14)	10(15/16)	
1											50
2											60
3											10
4											20
5											10
6											10
7											80
8											60
9											10
10											60
11											30
12											20
13											30
14											10

N = 10

$$\text{est } p = (0,5+0,6+0,1+0,2+0,1+0,1+0,8+0,6+0,1 + 0,6+0,3+0,2+0,3+0,1)/14 = .3286$$

GTadj = 5/14 = 0,2857

est p = .3286

Adj p = 0,5\*[(0,32857 - (1/10))\*(1-(1/10))]+0,5\*(0,3286/(1+0,2857))] = 0,2308



## Anhang G: Fragebögen zur Evaluation der minimalinvasiven Handgriffe (vgl. Kap. 4.2.2)

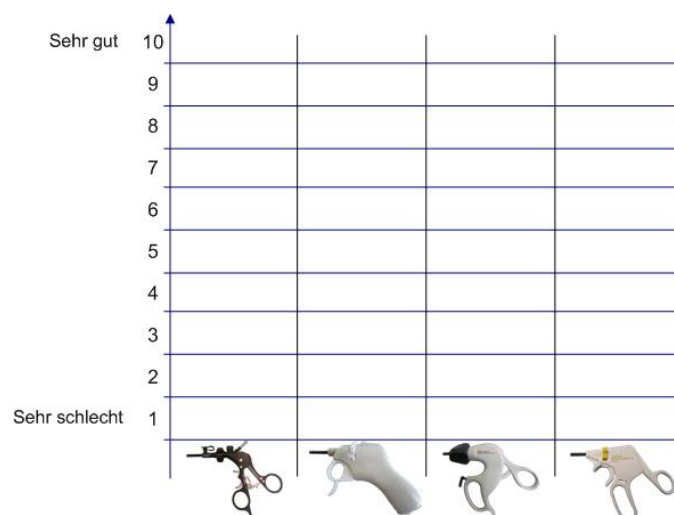
### ASQ:

<b>Die zuletzt gestellte Aufgabe konnte ich mit dem Handgriff auf einfache Weise durchführen!</b>	
Stimmt gar nicht <input type="checkbox"/> -3 <input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	Stimmt völlig <input type="checkbox"/> N/A <input type="checkbox"/> Nicht anwendbar
<b>Ich bin mit der Bearbeitungszeit für die letzte Aufgabe zufrieden!</b>	
Stimmt gar nicht <input type="checkbox"/> -3 <input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	Stimmt völlig <input type="checkbox"/> N/A <input type="checkbox"/> Nicht anwendbar
<b>Das Instrument hat mich in zufrieden stellender Weise bei der Erfüllung meiner Aufgabe unterstützt!</b>	
Stimmt gar nicht <input type="checkbox"/> -3 <input type="checkbox"/> -2 <input type="checkbox"/> -1 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	Stimmt völlig <input type="checkbox"/> N/A <input type="checkbox"/> Nicht anwendbar

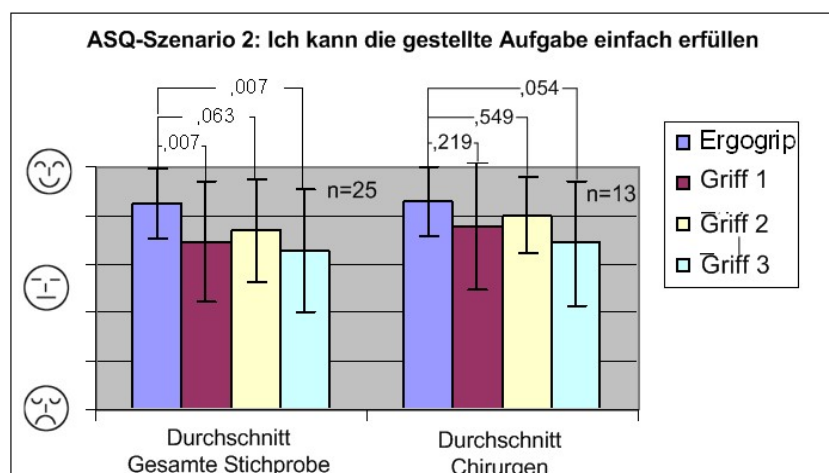
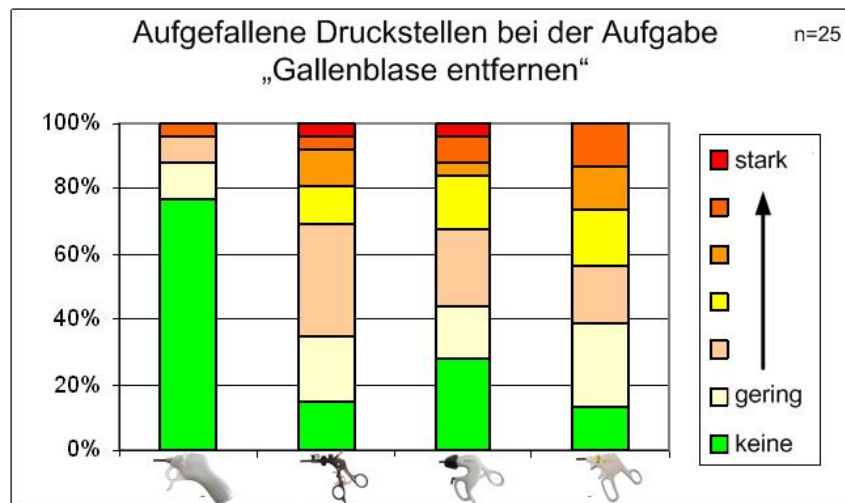
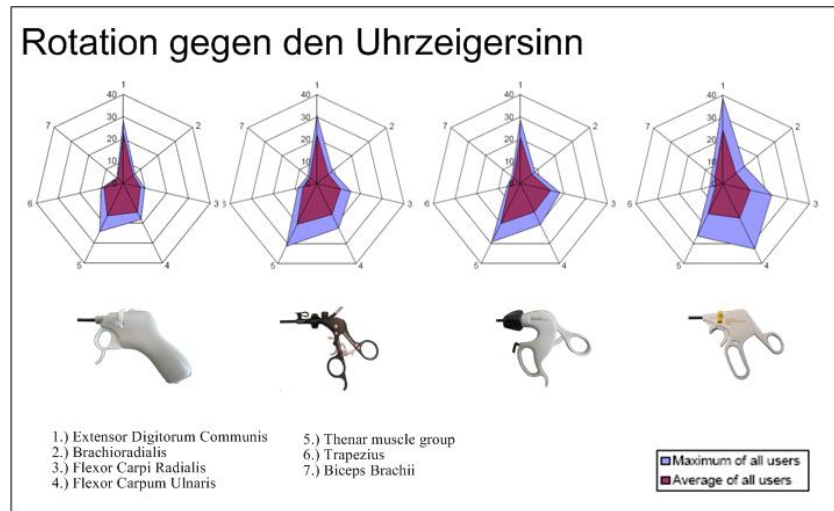
### Einschätzung von Schmerzen, Druckstellen und Verspannungen:

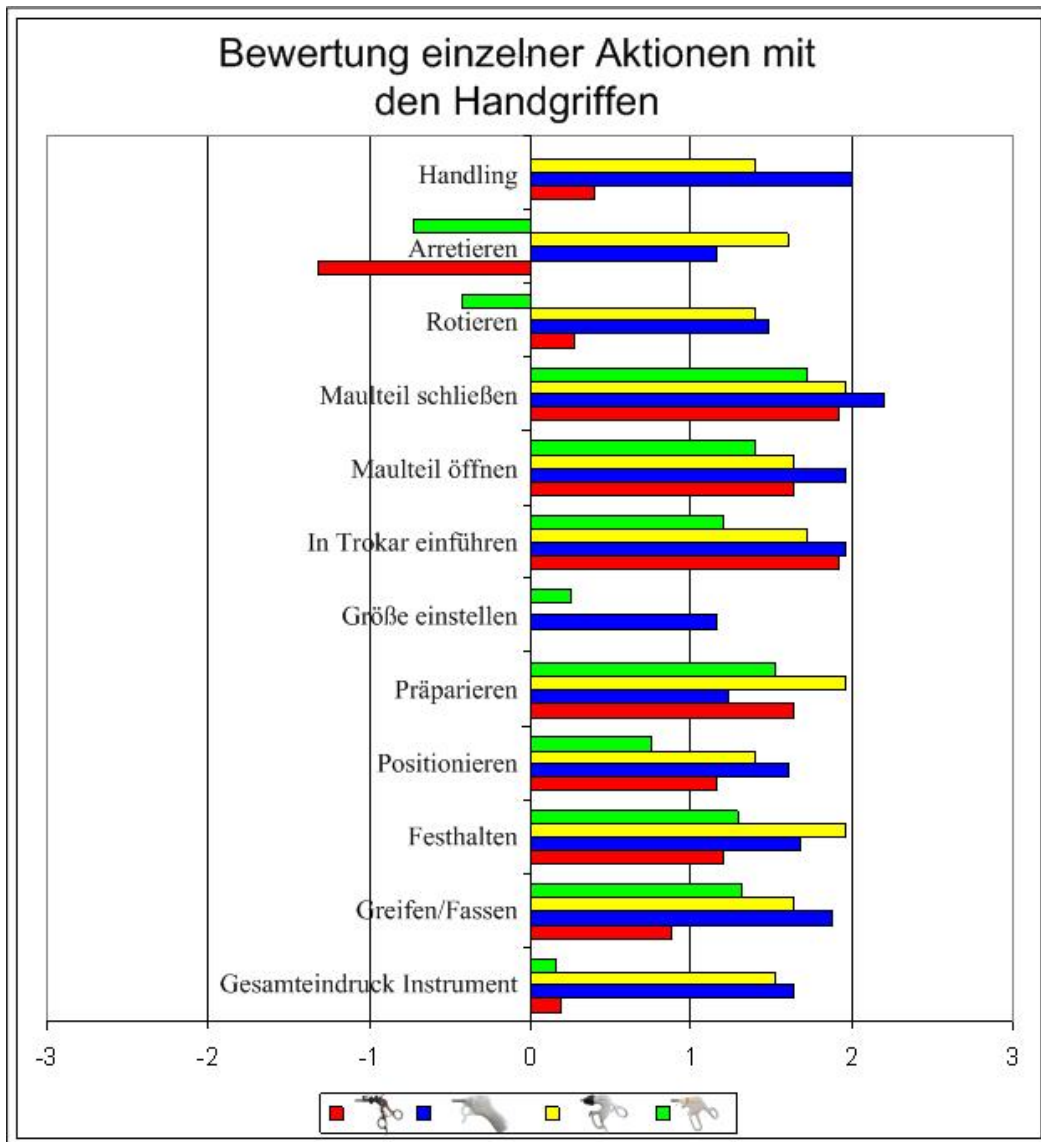
<b>Verspannungen:</b>	
gering <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 stark	Position: <input style="width: 100%;" type="text"/>
<b>Druckstellen:</b>	
gering <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 stark	Position: <input style="width: 100%;" type="text"/>
<b>Schmerzen:</b>	
gering <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 stark	Position: <input style="width: 100%;" type="text"/>

### Ranking:

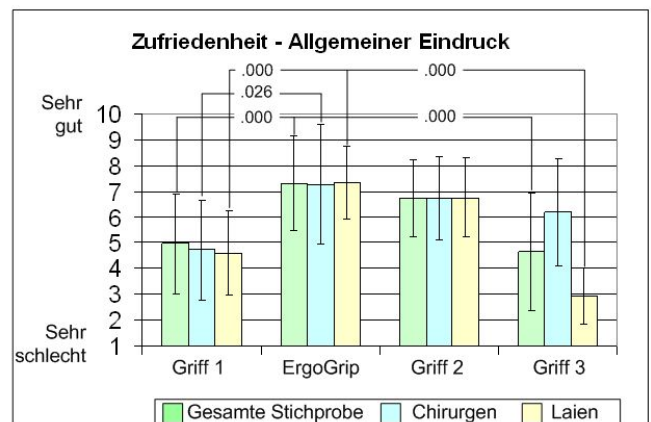
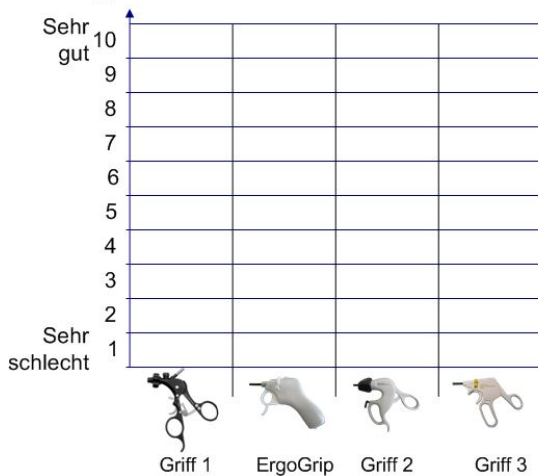


**Anhang H: Beispiele verschiedener Ergebnisse zur Messung der Gebrauchstauglichkeit der MIC-Handgriffe (vgl. Kap. 4.2.2)**  
EMG-Messung:





Beurteilen Sie die Handgriffe im Hinblick auf ihren allgemeinen Eindruck.



---

## Anhang I: Aufgaben zur Gebrauchstauglichkeitsevaluation einer Spritzenpumpe (vgl. Kap. 4.2.3 )

- 1.) Externe Einheit aus 4 Spritzenpumpen bauen
- 2.) Pumpenständer mit 3 Spritzenpumpen bestücken
- 3.) Spritzenpumpe einschalten
- 4.) Spritzenpumpe mit Spritze bestücken
- 6.) Rate von 10 ml/h eingeben, Förderung starten
- 7.) Bolus von 6 ml geben
- 8.) Spritze wechseln
- 9.) Dosierung eingeben: Volumen = 30ml, Dauer = 4 Stunden und Förderung starten
- 10.) Kontrast und Display-/Tastaturbeleuchtung einstellen
- 11.) Spritze wechseln
- 12.) Vorherige Therapie starten
- 13.) Dosis bei laufender Förderung um 5 ml/h erhöhen
- 14.) Bolusgabe von 7 ml in 14 Sekunden, danach mit vorheriger Dosierung fortfahren
- 15.) Laufende Förderung stoppen
- 16.) Standard-Bolusgabe auf 600 ml/h ändern
- 17.) Verabreichtes Gesamtvolumen notieren
- 18.) Standarddruck auf 797 mmHg einstellen
- 19.) Spritzenpumpe ausschalten

## Anhang J: Elemente der Problembewertung der zweiten Entwicklungsstudie (vgl. Kap. 6.2.2)

UseProb Bewertungstool 1.0

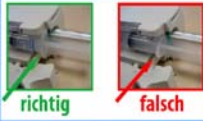
**2. Szenario: Spritzenpumpe**


Produktbeschreibung: Spritzenpumpe  
 Probanden: Anästhesietechnische Assistenten  
 Arbeitsaufgabe: Bestücken Sie den betriebsbereiten Perfusor mit der Spritze Omnilix 20ml, wählen Sie eine Rate von 10ml/h aus und starten Sie die Förderung

Problemszenario: Wird die Spritze "Omnilix 20ml" waagrecht anstatt senkrecht eingelegt, so kann im Menü nur der Sprizentyp "Omnilix 30ml" ausgewählt werden. Der Proband wählte den falschen Sprizentyp aus und stellte 10ml/h ein. Aufgrund des falschen Sprizentyps wurden nur 7,5ml/h injiziert.

Folge: Medikamentendosis war 25% geringer als gewünscht

Auftretenshäufigkeit: 1 Proband von 9 (11,11 %)



Beispielvideo: 

9 / 61

Szenarienbeschreibung

UseProb Bewertungstool 1.0

**2. Szenario: Spritzenpumpe**

Produktbeschreibung: Spritzenpumpe  
 Probanden: Anästhesietechnische Assistenten  
 Arbeitsaufgabe: Bestücken Sie den betriebsbereiten Perfusor mit der Spritze Omnilix 20ml, wählen Sie eine Rate von 10ml/h aus und starten Sie die Förderung

Problemszenario: Wird die Spritze "Omnilix 20ml" waagrecht anstatt senkrecht eingelegt, so kann im Menü nur der Sprizentyp "Omnilix 30ml" ausgewählt werden. Der Proband wählte den falschen Sprizentyp aus und stellte 10ml/h ein. Aufgrund des falschen Sprizentyps wurden nur 7,5ml/h injiziert.

Folge: Medikamentendosis war 25% geringer als gewünscht

Auftretenshäufigkeit: 1 Proband von 9 (11,11 %)

Auftretenshäufigkeit:  
 Ist die gestellte Aufgabe für Sie eine Primär- oder Sekundäraufgabe?  
 Primäraufgabe     Sekundäraufgabe  
 Wie oft wird Ihrer Meinung nach diese Aufgabe verwendet?  
 häufig     selten

10 / 61

Bewertung der Auftretenshäufigkeit

UseProb Bewertungstool 1.0

**2. Szenario: Spritzenpumpe**

Produktbeschreibung: Spritzenpumpe  
 Probanden: Anästhesietechnische Assistenten  
 Arbeitsaufgabe: Bestücken Sie den betriebsbereiten Perfusor mit der Spritze Omnilix 20ml, wählen Sie eine Rate von 10ml/h aus und starten Sie die Förderung

Problemszenario: Wird die Spritze "Omnilix 20ml" waagrecht anstatt senkrecht eingelegt, so kann im Menü nur der Sprizentyp "Omnilix 30ml" ausgewählt werden. Der Proband wählte den falschen Sprizentyp aus und stellte 10ml/h ein. Aufgrund des falschen Sprizentyps wurden nur 7,5ml/h injiziert.

Folge: Medikamentendosis war 25% geringer als gewünscht

Auftretenshäufigkeit: 1 Proband von 9 (11,11 %)

Bewertung der Aufgabenerfüllung:  
 Wie würden Sie die Aufgabenerfüllung bewerten?  
 Hervorragende Lösung     Mittleres Problem  
 Kein Problem     Großes Problem  
 Kosmetisches Problem     Katastrophe  
 Geringes Problem

11 / 61

Bewertung der Aufgabenbewältigung

UseProb Bewertungstool 1.0

**2. Szenario: Spritzenpumpe**

Produktbeschreibung: Spritzenpumpe  
 Probanden: Anästhesietechnische Assistenten  
 Arbeitsaufgabe: Bestücken Sie den betriebsbereiten Perfusor mit der Spritze Omnilix 20ml, wählen Sie eine Rate von 10ml/h aus und starten Sie die Förderung

Problemszenario: Wird die Spritze "Omnilix 20ml" waagrecht anstatt senkrecht eingelegt, so kann im Menü nur der Sprizentyp "Omnilix 30ml" ausgewählt werden. Der Proband wählte den falschen Sprizentyp aus und stellte 10ml/h ein. Aufgrund des falschen Sprizentyps wurden nur 7,5ml/h injiziert.

Folge: Medikamentendosis war 25% geringer als gewünscht

Auftretenshäufigkeit: 1 Proband von 9 (11,11 %)

Bitte nennen Sie die für Sie wahrscheinlichsten Folgen bei Eintritt des oben beschriebenen Problems.

Keine Auswirkung auf die Gesundheit des Patienten oder des Personals sowie keine Schäden am Material  
 Mögliche gefährliche Situation, die Sachschäden oder leichte Körperverletzungen zur Folge haben kann  
 Mögliche gefährliche Situation, die mittlere Körperverletzungen zur Folge haben kann  
 Mögliche gefährliche Situation, die schwere Körperverletzungen oder Tod zur Folge haben kann

12 / 61

Bewertung der Folgen des Problems

UseProb Bewertungstool 1.0

Medizinisches Gerät ist herausragend, Bedienbarkeit und Benutzerfreundlichkeit überragend positiv!

IX Medizinisches Gerät ist gut und erzeugt Zufriedenheit, die Erwartungen bezüglich Bedienbarkeit und Benutzerfreundlichkeit werden vorwiegend erfüllt!

VIII Medizinisches Gerät ist gut, Gebrauchstauglichkeitsprobleme sind sehr gering und treten selten auf. Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich!

VII Medizinisches Gerät ist befriedigend, irritiert in einzelnen Fällen, geringe Gebrauchstauglichkeitsprobleme sind festzustellen! Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich!

VI Medizinisches Gerät erfüllt die Erwartungen nicht vollständig, ist nur ausreichend

V Medizinisches Gerät ist nicht befriedigend, irritiert in einzelnen Fällen, geringe Gebrauchstauglichkeitsprobleme sind festzustellen! Personen- und Materialschäden aufgrund von Gebrauchstauglichkeitsproblemen sind höchst unwahrscheinlich!

IV Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsanforderungen! Durch Gebrauchstauglichkeitsprobleme verursachte gefährliche Situationen mit Sach- oder Personenschäden sind nicht auszuschließen, sollten die auftretenden Probleme nicht erkannt und behoben werden!

III Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsanforderungen! Es stellt ein Sicherheitsrisiko für den Patienten und/oder die Umgebung dar, sollten die auftretenden Probleme nicht erkannt und behoben werden!

II Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsanforderungen! Es stellt ein Sicherheitsrisiko für den Patienten und/oder die Umgebung dar, sollten die auftretenden Probleme nicht erkannt und behoben werden!

I Medizinisches Gerät hat durch Gebrauchstauglichkeitsprobleme bedingte Sicherheitsanforderungen! Es stellt ein Sicherheitsrisiko für den Patienten und/oder die Umgebung dar, sollten die auftretenden Probleme nicht erkannt und behoben werden!

13 / 61

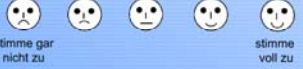
Ergebnis

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.


Wie stimmen Sie den folgenden Behauptungen zu?

Ich halte das Tool für nützlich, um die Gebrauchstauglichkeit in der Medizin zu bewerten.



stimme gar nicht zu    stimme voll zu

Ich halte derartige Tools allgemein als Bewertungshilfen für sinnvoll.



stimme gar nicht zu    stimme voll zu

54 / 61

Evaluation der Methode

## Anhang K: Fragebogen- Entwicklungsstudie 2

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

**Persönliche Angaben:**

Alter:

Geschlecht:  weibl.  männl.

Berufsgruppe:  Arzt  Ergonomie- / MMI-Experte  
 Pflege

Berufserfahrung:  Jahre

Wenn Sie damit einverstanden sind, dass Sie im Rahmen einer wissenschaftlichen Veröffentlichung evtl. namentlich zitiert werden, geben Sie bitte Ihren Namen in das folgende Feld ein:

52 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

**Evaluation Bewertungstool:**

1. Hat Sie das Tool bei der Bewertung der Probleme unterstützt?

gar nicht unterstützt voll unterstützt

2. Insgesamt fand ich die Bewertungsergebnisse ..

... genau richtig  
 ... zu positiv  
 ... zu negativ

53 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

**Wie stimmen Sie den folgenden Behauptungen zu?**

Ich halte das Tool für nützlich, um die Gebrauchstauglichkeit in der Medizin zu bewerten.

stimme gar nicht zu stimme voll zu

Ich halte derartige Tools allgemein als Bewertungshilfen für sinnvoll.

stimme gar nicht zu stimme voll zu

54 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

**Wie stimmen Sie der folgenden Behauptung zu?**

Es war für mich einfach das Tool zu verwenden.

stimme gar nicht zu stimme voll zu

55 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

**Was halten Sie von den Ergebniseinstufungen:**

Wie zufrieden sind Sie mit den gewählten Beschreibungen?

sehr unzufrieden sehr zufrieden

**Welcher Aussage stimmen Sie zu:**

Die Anzahl der Abstufungen wurde richtig gewählt.  
 Die Anzahl der Abstufungen ist zu hoch.  
 Die Anzahl der Abstufungen ist zu gering.

Mehr Infos hier!

56 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

Was finden Sie an dem Bewertungstool positiv:

57 / 61

UseProb Bewertungstool 1.0

Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

Was finden Sie an dem Bewertungstool negativ:

58 / 61

UseProb Bewertungstool 1.0


Bitte nehmen Sie sich noch kurz Zeit, einige Fragen zu beantworten.

Was würden Sie an dem Tool ändern:

59 / 61

## Anhang L: Bewertung der Szenarien bei der 2. Entwicklungsstudie (vgl. Kap. 6.2.2)

### Szenario 1

**UseProb Bewertungstool 1.0**


**1. Szenario: Praxissoftware**

Produktbeschreibung: Praxissoftware zur Rezepterstellung

Probanden: Ärzte

Arbeitsaufgabe: Bitte stellen Sie ein Rezept für die Medikamente: Acimethin Filmtabletten 25St (Harnwegsinfektion), Perenterol Forte 250 Mg 10 St. (Durchfall), Beloc-Zok 95 mg/rite 47,5 mg Retardtabletten (hoher Blutdruck) aus.

Problemszenario: Die Praxissoftware zeigt ein Rezeptformular mit drei Zeilen für zu verordnende Medikamente. Ein verordnetes Medikament wurde in Zeile 3 eingetragen. Aufgrund der langen Namen der Medikamente 1 und 2 bestand kein Platz mehr für den weiteren Ausdruck. Das dritte Medikament wurde somit nicht auf das Rezept gedruckt. Der Proband bemerkte dies nicht. Es erschien keine Fehlermeldung.

Folge: Die Aufgabe konnte nicht erfüllt werden. Das dritte Medikament stand nicht auf dem Rezept.

Auftretenshäufigkeit: 1 Proband von 12 (8,33%)

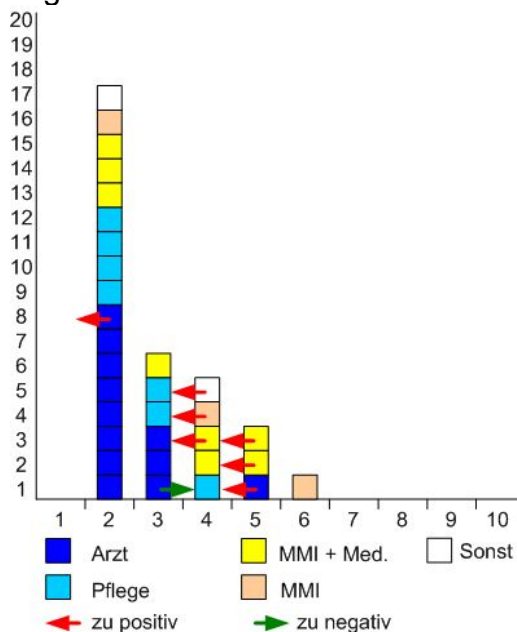
3 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	3,75	1,25	2,08	6,42	3,58	2,50
	Standardabweichung	,452	,452	,289	,669	,669	,905
	Minimum	3	1	2	5	2	2
	Maximum	4	2	3	7	4	5
Pflege	N	7	7	7	7	7	7
	Mittelwert	4,00	1,00	2,00	6,29	3,00	2,57
	Standardabweichung	,000	,000	,000	,756	1,000	,787
	Minimum	4	1	2	5	2	2
	Maximum	4	1	2	7	4	4
MMI	N	3	3	3	3	3	3
	Mittelwert	4,00	1,00	2,00	5,67	2,67	4,00
	Standardabweichung	,000	,000	,000	,577	1,528	2,000
	Minimum	4	1	2	5	1	2
	Maximum	4	1	2	6	4	6
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	3,75	1,25	2,00	6,00	3,00	3,38
	Standardabweichung	,463	,463	,000	,756	,756	1,302
	Minimum	3	1	2	5	2	2
	Maximum	4	2	2	7	4	5
Sonst	N	2	2	2	2	2	2
	Mittelwert	4,00	1,00	2,00	5,50	3,50	3,00
	Standardabweichung	,000	,000	,000	,707	,707	1,414
	Minimum	4	1	2	5	3	2
	Maximum	4	1	2	6	4	4
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	3,84	1,16	2,03	6,16	3,22	2,91
	Standardabweichung	,369	,369	,177	,723	,870	1,174
	Minimum	3	1	2	5	1	2
	Maximum	4	2	3	7	4	6



Bei diesem Szenario liegt der Mittelwert der Bewertung über die gesamte Stichprobe bei 2,91 mit einer Standardabweichung von 1,17.

Abb. A veranschaulicht die Meinungsverteilung der einzelnen Evaluatoren, farblich nach Fachwissen aufgeteilt.



**Abb. A: Bewertung von Szenario 1**

24 von 32 Testpersonen waren mit ihrer Bewertung zufrieden (75%). Diejenigen, welche mit ihrer Bewertung nicht zufrieden waren, fanden das Ergebnis mit Ausnahme einer Person zu positiv. Hiervon lagen fast alle mit ihrem Ergebnis über dem Durchschnitt (Abb. A). Da die meisten das Ergebnis als genau richtig einstufen, kann davon ausgegangen werden, dass für dieses Szenario die Stufeneinteilung richtig gewählt wurde. Für die unzufriedenen Anwender könnte in Betracht gezogen werden, das Bewertungsergebnis bei einer unpassenden Bewertung entsprechend ihrer Präferenzen zu adaptieren. Viele der Evaluatoren, welche durch ihre Einstufung sehr vom Mittelwert abwichen, hätten sich durch eine derartige Option dem Mittelwert angenähert.

Bei diesem Szenario zeigte sich, dass die Evaluatoren mit MMI-Erfahrung das Problem besser bewerteten. Die im Vergleich zur medizinischen Stichprobe höhere Standardabweichung gibt zusätzlich zu erkennen, dass die Nutzer bei der Bewertung weiter voneinander abwichen. Außerdem wurde in diesem Szenario deutlich, dass das medizinische Personal bei der Bewertung gleichgesinnter war. Die schlechtere Notenvergabe kam dadurch zustande, dass diese Personengruppe in der Regel die Folgen der Probleme als schwerwiegender bewertete.

## Szenario 2

**UseProb Bewertungstool 1.0**

 UNIVERSITÄT  
TÜBINGEN

**2. Szenario: Spritzenpumpe**

Produktbeschreibung: Spritzenpumpe  
 Probanden: Anästhesietechnische Assistenten  
 Arbeitsaufgabe: Bestücken Sie den betriebsbereiten Perfusor mit der Spritze Omnifix 20ml, wählen Sie eine Rate von 10ml/h aus und starten Sie die Förderung.

Problemszenario: Wird die Spritze "Omnifix 20ml" waagrecht anstatt senkrecht eingelegt, so kann im Menü nur der Sprizentyp "Omnifix 30ml" ausgewählt werden. Der Proband wählte den falschen Sprizentyp aus und stellte 10ml/h ein. Aufgrund des falschen Sprizentyps wurden nur 7,5ml/h injiziert.

Folge: Medikamentendosis war 25% geringer als gewünscht  
 Auftretenshäufigkeit: 1 Proband von 9 (11,11 %)



**richtig**



**falsch**

Beispielvideo: 

9 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	4,00	1,00	2,00	6,67	3,92	2,00
	Standardabweichung	,000	,000	,000	,492	,289	,000
	Minimum	4	1	2	6	3	2
	Maximum	4	1	2	7	4	2
Pflege	N	7	7	7	7	7	7
	Mittelwert	4,00	1,00	2,00	6,71	3,86	2,00
	Standardabweichung	,000	,000	,000	,488	,378	,000
	Minimum	4	1	2	6	3	2
	Maximum	4	1	2	7	4	2
MMI	N	3	3	3	3	3	3
	Mittelwert	4,00	1,00	2,00	6,00	3,33	2,67
	Standardabweichung	,000	,000	,000	1,000	,577	1,155
	Minimum	4	1	2	5	3	2
	Maximum	4	1	2	7	4	4
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	4,00	1,00	2,00	5,88	3,50	2,50
	Standardabweichung	,000	,000	,000	,641	,756	,756
	Minimum	4	1	2	5	2	2
	Maximum	4	1	2	7	4	4
Sonst	N	2	2	2	2	2	2
	Mittelwert	4,00	1,00	2,00	6,50	3,50	2,00
	Standardabweichung	,000	,000	,000	,707	,707	,000
	Minimum	4	1	2	6	3	2
	Maximum	4	1	2	7	4	2
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	4,00	1,00	2,00	6,41	3,72	2,19
	Standardabweichung	,000	,000	,000	,665	,523	,535
	Minimum	4	1	2	5	2	2
	Maximum	4	1	2	7	4	4

Beim zweiten Szenario liegt der Mittelwert der Bewertung bei 2,19 mit einer Standardabweichung von 0,535.

28 von 32 Testpersonen waren mit ihrer Bewertung zufrieden (87,5%). Drei Ärzte fanden die Bewertung mit „2“ zu positiv, jedoch war eine Bewertung mit der negativeren Note „1“ nicht möglich, aufgrund der geringen Prozentzahl, mit der das

Problem auftrat (vgl. Kap. 6.1.1). Insgesamt war sich die Stichprobe bei der Bewertung einig. Lediglich vier der Befragten wichen von der allgemeinen Bewertung ab (Abb. B). Bei den Probanden mit medizinischem Fachwissen (Ärzte und Pflegekräfte) war keine Abweichung zu erkennen. Im Mittelwert stuften die Evaluatoren mit MMI-Erfahrung das Problem besser ein.

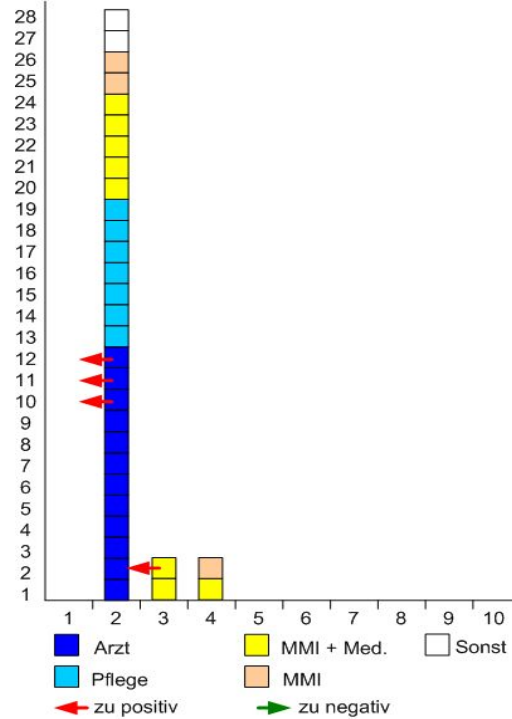


Abb. B: Bewertung von Szenario 2

Szenario 3

**UseProb Bewertungstool 1.0**

**3. Szenario: HF-Gerät 1**

Produktbeschreibung: Hochfrequenz-Chirurgiegerät

Probanden: Krankenschwestern

Arbeitsaufgabe: Bitte erhöhen Sie die Leistung von 50 auf 60 Watt

Problemszenario: Die für die Einstellung notwendigen Tasten (in Abb. "rot") werden nicht erkannt. Es wird zuerst mit Hilfe der dem Display nahegelegenen Tasten ("gelb") versucht, die Leistung zu verändern. Dies führt jedoch zu keiner Veränderung. Nach einer gewissen Zeit können alle Probanden die Leistung einstellen.

Folge: Die Aufgabe konnte erfüllt werden. Geringe Beeinträchtigung der Zufriedenheit der Probanden war feststellbar. Das Problem verursachte einen durchschnittlichen zeitlichen Mehraufwand von 22 Sekunden.

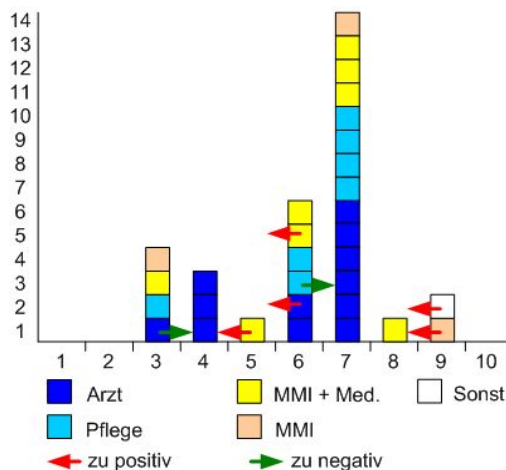
Auftretenshäufigkeit: 5 Probanden von 16 (31,25 %)

Beispielvideo:

15 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	5,00	1,42	2,00	4,42	1,50	5,75
	Standardabweichung	,000	,515	,000	,515	,674	1,545
	Minimum	5	1	2	4	1	3
	Maximum	5	2	2	5	3	7
Pflege	N	7	7	7	7	7	7
	Mittelwert	5,00	1,14	1,86	4,43	1,29	6,14
	Standardabweichung	,000	,378	,378	,787	,488	1,464
	Minimum	5	1	1	4	1	3
	Maximum	5	2	2	6	2	7
MMI	N	3	3	3	3	3	3
	Mittelwert	4,33	1,67	1,33	4,33	1,33	6,33
	Standardabweichung	,577	,577	,577	1,528	,577	3,055
	Minimum	4	1	1	3	1	3
	Maximum	5	2	2	6	2	9
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	4,88	1,25	1,88	4,13	1,75	6,13
	Standardabweichung	,354	,463	,354	,641	,886	1,553
	Minimum	4	1	1	3	1	3
	Maximum	5	2	2	5	3	8
Sonst	N	1	1	1	1	1	1
	Mittelwert	4,00	2,00	1,00	3,00	1,00	9,00
	Standardabweichung	.	.	.	.	.	.
	Minimum	4	2	1	3	1	9
	Maximum	4	2	1	3	1	9
Insgesamt	N	31	31	31	31	31	31
	Mittelwert	4,87	1,35	1,84	4,29	1,48	6,10
	Standardabweichung	,341	,486	,374	,739	,677	1,680
	Minimum	4	1	1	3	1	3
	Maximum	5	2	2	6	3	9

Bei der Bewertung des dritten Szenarios liegt der Mittelwert für die gesamte Stichprobe bei 6,10 mit einer Standardabweichung von 1,68.



**Abb. C: Bewertung von Szenario 3**

Bei dieser Bewertung waren 24 von 31 Evaluatoren (77,4%) mit dem Ergebnis zufrieden (Abb. C). Ein Evaluator lieferte aufgrund eines technischen Problems des Bewertungstools kein Ergebnis. Bis auf zwei Ärzte und eine Pflegekraft waren die Gruppen der realen Nutzer von Medizinprodukten mit dem Bewertungsergebnis zufrieden. Die relativ hohe Standardabweichung zeigt, dass das Problem unterschiedlich eingeschätzt wurde. Man erkennt, dass dieses Problem die Stichprobe in zwei Gruppen teilt. Obwohl es unterschiedliche Bewertungsergebnisse gibt, sind die meisten Evaluatoren mit ihrem Ergebnis zufrieden, ein Zeichen für die Existenz unterschiedlicher Meinungen in dieser Bewertung.

## Szenario 4

**UseProb Bewertungstool 1.0**

**4. Szenario: HF-Gerät 2**

Produktbeschreibung: Hochfrequenz-Chirurgiegerät  
 Probanden: Krankenschwestern  
 Arbeitsaufgabe: Bitte schließen Sie einen Fußschalter an das Gerät an und führen Sie eine Koagulation mit Hilfe des Fusschalters durch.

Problemszenario: Das HF-Gerät bietet die Möglichkeit, drei Fußschalter anzuschließen. Buchse 1 und Buchse 3 besitzen denselben Steckeranschluss. Buchse 2 ist verplombt. Der Fußschalter wurde an Buchse 1 angeschlossen. Dies führte dazu, dass bei der Erledigung der Aufgaben der Fußschalter umgesteckt werden musste, da die gewünschte Fußschalterfunktion nicht, der Buchse 1 zugewiesen werden konnte.

Folge: Die Aufgabe konnte von allen Probanden erfüllt werden. Die Probanden waren unzufrieden, und es entstand ein durchschnittlicher Zeitverlust von 4 min. 20sek.

Auftretenshäufigkeit: 12 Probanden von 16 (75 %)

Beispielvideo:

21 / 61

Gruppen	Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung	
Arzt	N	12	12	12	12	12	
	Mittelwert	4,92	1,42	1,75	5,00	1,83	4,67
	Standardabweichung	,289	,515	,452	,853	,937	2,103
	Minimum	4	1	1	4	1	1
	Maximum	5	2	2	6	4	7
Pflege	N	7	7	7	7	7	
	Mittelwert	5,00	1,00	2,00	4,71	1,57	5,29
	Standardabweichung	,000	,000	,000	,488	,535	1,254
	Minimum	5	1	2	4	1	4
	Maximum	5	1	2	5	2	7
MMI	N	3	3	3	3	3	
	Mittelwert	4,67	1,67	1,67	5,33	2,00	3,67
	Standardabweichung	,577	,577	,577	,577	,000	,577
	Minimum	4	1	1	5	2	3
	Maximum	5	2	2	6	2	4
MMI+Med	N	8	8	8	8	8	
	Mittelwert	4,88	1,63	1,88	5,38	1,75	4,75
	Standardabweichung	,354	,518	,354	,518	,886	1,669
	Minimum	4	1	1	5	1	3
	Maximum	5	2	2	6	3	7
Sonst	N	2	2	2	2	2	
	Mittelwert	5,00	1,00	2,00	5,00	2,50	4,00
	Standardabweichung	,000	,000	,000	1,414	2,121	4,243
	Minimum	5	1	2	4	1	1
	Maximum	5	1	2	6	4	7
Insgesamt	N	32	32	32	32	32	
	Mittelwert	4,91	1,38	1,84	5,06	1,81	4,69
	Standardabweichung	,296	,492	,369	,716	,859	1,822
	Minimum	4	1	1	4	1	1
	Maximum	5	2	2	6	4	7

Das Szenario 4 (Abb. D) bewertete die Stichprobe im Mittel mit 4,69 mit einer Standardabweichung von 1,82.

Bei dieser Bewertung waren 27 von 32 Evaluatoren (84,38%) mit dem Ergebnis zufrieden (Abb. D). Dennoch war sich die Stichprobe uneinig. Dieses Szenario

polarisierte. Dies lag größtenteils an der Einschätzung der Auswirkung des Problems. Auch bei dieser Szenarienbewertung wird erkenntlich, dass Meinungen bei der Bewertung voneinander abweichen.

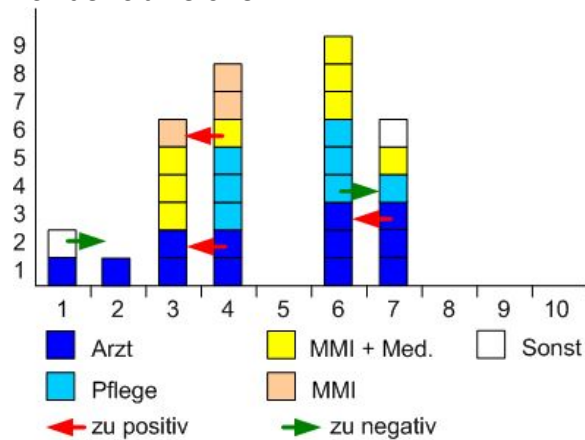


Abb. D: Bewertung von Szenario 4

### Szenario 5

UseProb Bewertungstool 1.0




**5. Szenario: Medikamentendosis**

Produktbeschreibung: Medikamentenverordnungssystem

Probanden: Krankenschwestern

Arbeitsaufgabe: Bitte ändern Sie die Dosis des zu verabreichenden Medikamentes von 10 mg/Tag auf 15mg/Tag.

Problemszenario: Die Probanden gaben die neue Dosis ein, ohne die vorherige Dosis zu löschen. In der Folge erhielten die Patienten die Summe aus alter und neuer Dosis. Die Probanden erhielten keine Nachricht, welche sie auf diese Doppelung aufmerksam gemacht hätte.

Folge: Der Patient erhält die Summe aus alter und neuer Dosis.

Auftretenshäufigkeit: 4 Probanden von 19 (21 %)

27 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	4,83	1,17	2,00	6,58	3,83	1,42
	Standardabweichung	,389	,389	,000	,669	,389	,669
	Minimum	4	1	2	5	3	1
	Maximum	5	2	2	7	4	3
Pflege	N	7	7	7	7	7	7
	Mittelwert	4,86	1,14	2,00	6,71	3,71	1,14
	Standardabweichung	,378	,378	,000	,488	,488	,378
	Minimum	4	1	2	6	3	1
	Maximum	5	2	2	7	4	2
MMI	N	3	3	3	3	3	3
	Mittelwert	5,00	1,00	2,00	6,67	4,00	1,00
	Standardabweichung	,000	,000	,000	,577	,000	,000
	Minimum	5	1	2	6	4	1
	Maximum	5	1	2	7	4	1
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	5,00	1,00	2,00	6,75	3,75	1,25
	Standardabweichung	,000	,000	,000	,463	,463	,463
	Minimum	5	1	2	6	3	1
	Maximum	5	1	2	7	4	2
Sonst	N	2	2	2	2	2	2
	Mittelwert	4,50	1,00	2,00	4,50	3,50	4,00
	Standardabweichung	,707	,000	,000	,707	,707	1,414
	Minimum	4	1	2	4	3	3
	Maximum	5	1	2	5	4	5
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	4,88	1,09	2,00	6,53	3,78	1,44
	Standardabweichung	,336	,296	,000	,761	,420	,878
	Minimum	4	1	2	4	3	1
	Maximum	5	2	2	7	4	5

Der Mittelwert über die gesamte Stichprobe bei Szenario 5 lag bei 1,44 mit einer Standardabweichung von 0,878.

In diesem Szenario waren sich die Befragten sehr einig. 28 von 32 Teilnehmern (87,5%) waren mit dem Bewertungsergebnis zufrieden (Abb. E), was auch durch die relativ geringe Standardabweichung erkenntlich wird. Drei der Unzufriedenen würden ihren Wert näher zum Mittelwert korrigieren.

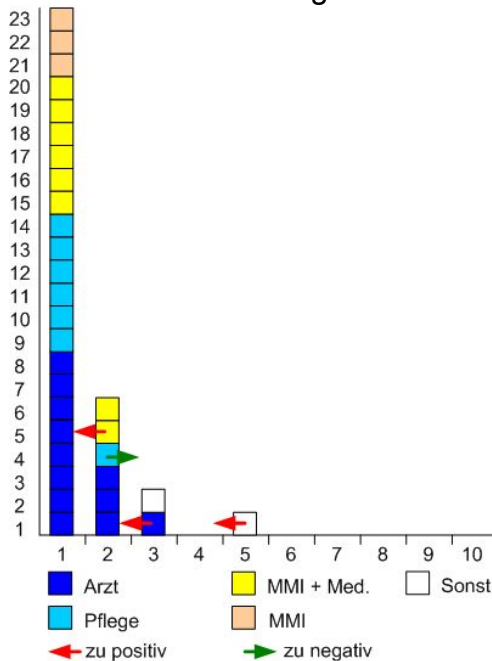


Abb. E: Bewertung von Szenario 5

## Szenario 6

**UseProb Bewertungstool 1.0**

**6. Szenario: Monitorkontrast**

**Produktbeschreibung:** OP-Monitor  
**Probanden:** Krankenschwestern  
**Arbeitsaufgabe:** Bitte passen Sie den Kontrast und die Helligkeit des Monitors an die gegebenen Lichtverhältnisse an.

**Problemszenario:** Die Probanden hatten Schwierigkeiten, das Einstellungs Menü für Kontrast und Helligkeit zu finden. Viele verirrt sich in anderen Menüebenen und konnten erst durch einen Anruf beim Service oder mit Hilfe des Benutzerhandbuches die Aufgabe lösen.

**Folge:** Die Aufgabe konnte erfüllt werden. Die Probanden waren jedoch unzufrieden, und es entstand ein durchschnittlicher Zeitverlust von 3 min. 12 sek.

**Auftretenshäufigkeit:** 12 Probanden von 20 (60 %)

33 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	5,75	1,92	1,67	4,58	1,50	5,58
	Standardabweichung	,622	,289	,492	,793	,798	2,021
	Minimum	5	1	1	4	1	2
	Maximum	7	2	2	6	3	7
Pflege	N	7	7	7	7	7	7
	Mittelwert	6,14	1,43	1,57	4,43	1,57	5,57
	Standardabweichung	1,069	,535	,535	1,134	,976	2,573
	Minimum	5	1	1	3	1	2
	Maximum	7	2	2	6	3	8
MMI	N	3	3	3	3	3	3
	Mittelwert	5,67	1,67	1,33	4,00	1,00	7,00
	Standardabweichung	1,155	,577	,577	1,000	,000	1,000
	Minimum	5	1	1	3	1	6
	Maximum	7	2	2	5	1	8
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	5,50	1,75	1,13	4,63	1,13	6,13
	Standardabweichung	,926	,463	,354	,916	,354	1,458
	Minimum	5	1	1	3	1	3
	Maximum	7	2	2	6	2	8
Sonst	N	2	2	2	2	2	2
	Mittelwert	5,00	2,00	1,00	3,00	1,00	8,00
	Standardabweichung	,000	,000	,000	1,414	,000	1,414
	Minimum	5	2	1	2	1	7
	Maximum	5	2	1	4	1	9
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	5,72	1,75	1,44	4,41	1,34	6,00
	Standardabweichung	,851	,440	,504	,979	,701	1,951
	Minimum	5	1	1	2	1	2
	Maximum	7	2	2	6	3	9

Bei diesem Szenario liegt der Mittelwert bei 6,00 mit einer Standardabweichung von 1,951.

23 von 32 Testpersonen (71,88%) waren mit ihrer Bewertung zufrieden. In diesem Szenario war die Standardabweichung relativ hoch. Gerade die Anwender in den



Pflegeberufen differierten sehr bei der Einstufung der Folgen (Abb. F). So entstanden zwei Gruppen: eine kleinere, die das Problem als sicherheitskritisch beurteilte, und eine größere, welche es als nicht sicherheitskritisch einstufte.

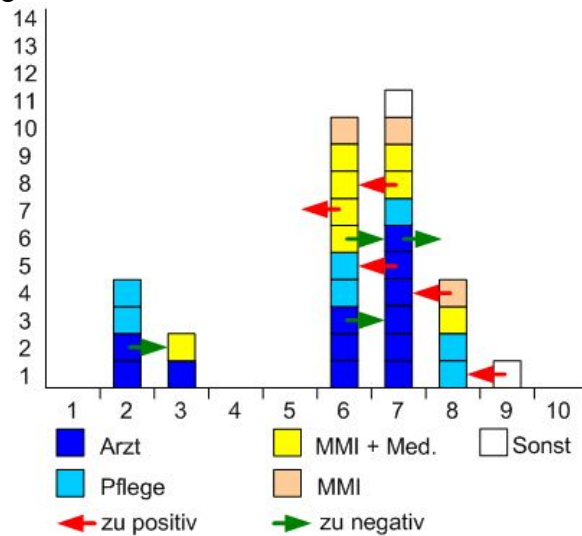


Abb. F: Bewertung von Szenario 6

### Szenario 7

UseProb Bewertungstool 1.0

SARRENSBERG  
UNIVERSITÄT  
TÜBINGEN

**7. Szenario: Hinweismeldung**

**Produktbeschreibung:** Hochfrequenz-Chirurgiegerät

**Probanden:** Krankenschwestern

**Arbeitsaufgabe:** Das Problem tritt auf, wenn kein Modus für Koagulieren oder Schneiden ausgewählt ist, und versucht wird, zu koagulieren oder zu schneiden.

**Problemszenario:** Hinweismeldungen werden auf hellgrünem Hintergrund mit schwarzer Schrift angegeben.

**Folge:** Die Kombination ist schlecht lesbar, und der grüne Hintergrund kann zu einer falschen Interpretation führen, da er Sicherheit suggeriert. Probleme innerhalb des Usability Tests traten nicht auf.

**Auftretenshäufigkeit:** 0 Probanden von 16 (0 %). Das Problem fiel jedoch in der Expertenbegutachtung auf.

39 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	2,33	1,67	1,50	3,67	1,33	8,25
	Standardabweichung	,492	,492	,522	,651	,651	,965
	Minimum	2	1	1	3	1	6
	Maximum	3	2	2	5	3	9
Pflege	N	7	7	7	7	7	7
	Mittelwert	2,57	1,43	1,71	4,29	1,29	7,29
	Standardabweichung	,535	,535	,488	,951	,488	1,704
	Minimum	2	1	1	3	1	5
	Maximum	3	2	2	5	2	9
MMI	N	3	3	3	3	3	3
	Mittelwert	3,00	1,00	2,00	3,67	1,00	8,33
	Standardabweichung	,000	,000	,000	1,155	,000	1,155
	Minimum	3	1	2	3	1	7
	Maximum	3	1	2	5	1	9
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	2,75	1,25	1,50	4,13	1,75	7,13
	Standardabweichung	,463	,463	,535	1,126	,886	1,885
	Minimum	2	1	1	3	1	5
	Maximum	3	2	2	6	3	9
Sonst	N	2	2	2	2	2	2
	Mittelwert	2,00	2,00	1,00	3,00	1,00	9,00
	Standardabweichung	,000	,000	,000	,000	,000	,000
	Minimum	2	2	1	3	1	9
	Maximum	2	2	1	3	1	9
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	2,53	1,47	1,56	3,88	1,38	7,81
	Standardabweichung	,507	,507	,504	,907	,660	1,469
	Minimum	2	1	1	3	1	5
	Maximum	3	2	2	6	3	9

Bei dem siebten Szenario (Abb. G) liegt der Mittelwert bei 7,81 mit einer Standardabweichung von 1,47.

21 von 32 Testpersonen (65,63%) waren mit ihrer Bewertung zufrieden. Alle mit dem Ergebnis unzufriedenen Probanden empfanden das ermittelte Ergebnis als zu positiv. Diese positive Beurteilung entstand hauptsächlich dadurch, dass die Auftretenshäufigkeit gering war, weil kein Proband mit dem Problem konfrontiert wurde. Hierbei stellt sich die Frage, ob die Stufen für die Auftretenshäufigkeit angepasst werden müssen, oder ob es ausreicht, dass unzufriedene Anwender der Methode am Ende ihre Bewertungsergebnisse adaptieren können.

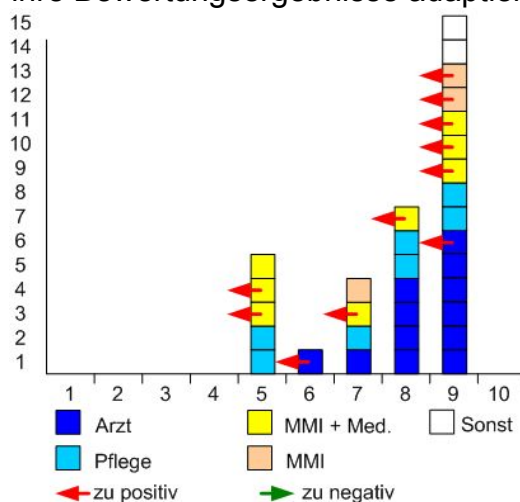


Abb. G: Bewertung von Szenario 7

## Szenario 8

**UseProb Bewertungstool 1.0**

**8. Szenario: "Press any Key"**

**Produktbeschreibung:** Hochfrequenz-Chirurgiegerät  
**Probanden:** Krankenschwestern  
**Arbeitsaufgabe:** Bitte machen Sie das HF-Gerät betriebsbereit.  
**Problemszenario:** Es erscheint eine Meldung: "Systemtest - Press Any Key". Die Betätigung der Instrumententasten am Handgriff wird nicht als Bestätigung erkannt. Es muss eine Taste am Gerät gedrückt werden.  
**Folge:** Die Probanden verwendeten die Instrumententasten als "any key". Es passierte nichts. Sie erkannten das Problem und drückten eine Taste am Gerät. Die Probanden waren verunsichert, es entstand ein durchschnittlicher Zeitverlust von 12 Sekunden.  
**Auftretenshäufigkeit:** 12 Probanden von 16 (75 %)

Program:

System-Test  
Press any key

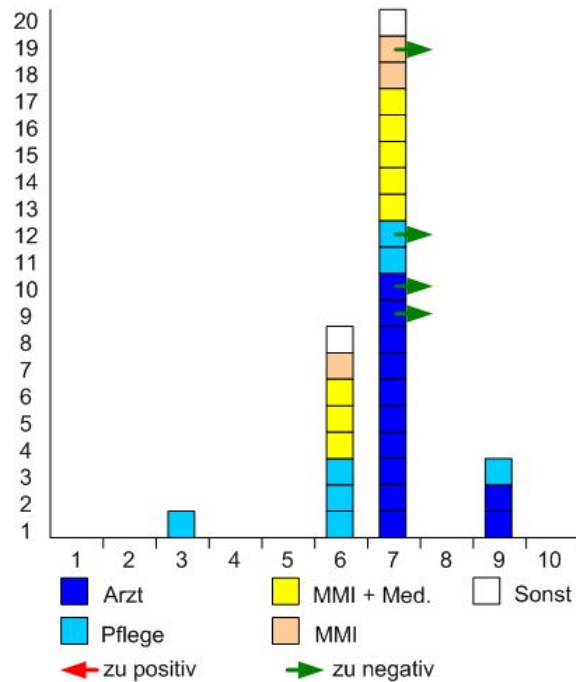
Intro:

45 / 61

Gruppen		Auftreten	Aufgabe	Häufigkeit	Problem	Folgen	Auswertung
Arzt	N	12	12	12	12	12	12
	Mittelwert	6,75	1,33	1,92	3,33	1,00	7,33
	Standardabweichung	,452	,492	,289	,651	,000	,778
	Minimum	6	1	1	2	1	7
	Maximum	7	2	2	4	1	9
Pflege	N	7	7	7	7	7	7
	Mittelwert	6,86	1,14	2,00	4,00	1,14	6,29
	Standardabweichung	,378	,378	,000	1,291	,378	1,799
	Minimum	6	1	2	2	1	3
	Maximum	7	2	2	5	2	9
MMI	N	3	3	3	3	3	3
	Mittelwert	7,00	1,00	2,00	4,00	1,00	6,67
	Standardabweichung	,000	,000	,000	1,000	,000	,577
	Minimum	7	1	2	3	1	6
	Maximum	7	1	2	5	1	7
MMI+Med	N	8	8	8	8	8	8
	Mittelwert	6,50	1,38	1,88	4,38	1,00	6,63
	Standardabweichung	,756	,518	,354	,518	,000	,518
	Minimum	5	1	1	4	1	6
	Maximum	7	2	2	5	1	7
Sonst	N	2	2	2	2	2	2
	Mittelwert	7,00	1,00	2,00	4,50	1,00	6,50
	Standardabweichung	,000	,000	,000	,707	,000	,707
	Minimum	7	1	2	4	1	6
	Maximum	7	1	2	5	1	7
Insgesamt	N	32	32	32	32	32	32
	Mittelwert	6,75	1,25	1,94	3,88	1,03	6,81
	Standardabweichung	,508	,440	,246	,907	,177	1,061
	Minimum	5	1	1	2	1	3
	Maximum	7	2	2	5	2	9

Bei dem letzten bewerteten Szenario liegt der Mittelwert bei 6,81, mit einer Standardabweichung von 1,06.

Bei diesem Szenario waren 28 von 32 (87,5%) Evaluatoren mit dem Bewertungsergebnis zufrieden. Die meisten Bewertungen lagen eng beieinander. Lediglich vier waren in ihrer Abweichung auffallend, die Probanden waren jedoch mit ihrem Ergebnis zufrieden (Abb. H).



**Abb. H: Bewertung von Szenario 8**

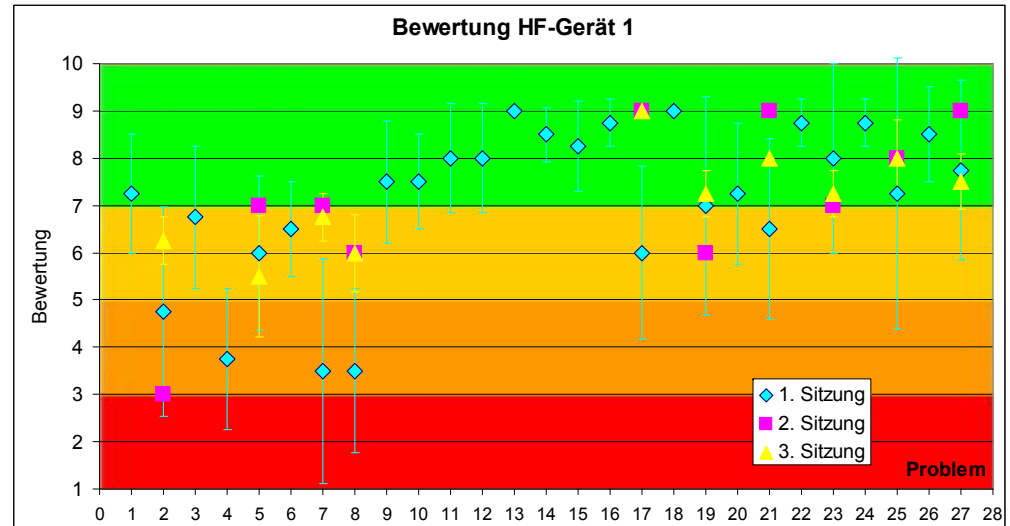
**Anhang M: Subjektive Bewertung der Fachexperten in der 2. Evaluationsstudie (vgl. Kap. 6.2.2)**

<p><b>Hat Sie das Tool bei der Bewertung unterstützt?</b></p> <p>Detailed description: A horizontal bar chart with five bars representing different categories. The x-axis is a Likert scale from 1 (happy face) to 5 (sad face). The bars are: MMI (score ~3.5), MMI+Med (score ~2.5), Pflege (score ~2.5), Arzt (score ~2.5), and Insgesamt (score ~2.5).</p>	<table border="0"> <tr> <td>Insgesamt: 1,69</td> <td>Standardabw.: 0,644</td> </tr> <tr> <td>Arzt: 1,67</td> <td>Standardabw.: 0,651</td> </tr> <tr> <td>Pflege: 1,43</td> <td>Standardabw.: 0,535</td> </tr> <tr> <td>MMI+Med.: 1,75</td> <td>Standardabw.: 0,707</td> </tr> <tr> <td>MMI.: 2,33</td> <td>Standardabw.: 0,577</td> </tr> <tr> <td>Sonst: 1,50</td> <td>Standardabw.: 0,707</td> </tr> </table>	Insgesamt: 1,69	Standardabw.: 0,644	Arzt: 1,67	Standardabw.: 0,651	Pflege: 1,43	Standardabw.: 0,535	MMI+Med.: 1,75	Standardabw.: 0,707	MMI.: 2,33	Standardabw.: 0,577	Sonst: 1,50	Standardabw.: 0,707
Insgesamt: 1,69	Standardabw.: 0,644												
Arzt: 1,67	Standardabw.: 0,651												
Pflege: 1,43	Standardabw.: 0,535												
MMI+Med.: 1,75	Standardabw.: 0,707												
MMI.: 2,33	Standardabw.: 0,577												
Sonst: 1,50	Standardabw.: 0,707												
<p><b>Insgesamt finde ich die Bewertungsergebnisse...</b></p> <p>Detailed description: A horizontal bar chart with three bars. The x-axis is labeled 'Probanden' and ranges from 0 to 30. The bars are: ...zu negativ (0), ...zu positiv (4), and ...genau richtig (28).</p>	<p>Insgesamt finde ich die Bewertungsergebnisse:</p> <p>Genau richtig: 28          Zu positiv: 4          (Sonst, Arzt, MMI+MED 2x)          Zu negativ: 0</p>												
<p><b>Ich halte das Tool für nützlich um die Gebrauchstauglichkeit in der Medizin zu bewerten!</b></p> <p>Detailed description: A horizontal bar chart with five bars representing different categories. The x-axis is a Likert scale from 1 (happy face) to 5 (sad face). The bars are: MMI (score ~3.5), MMI+Med (score ~2.5), Pflege (score ~2.5), Arzt (score ~2.5), and Insgesamt (score ~2.5).</p>	<p>Ich halte das Tool für nützlich, um die Gebrauchstauglichkeit in der Medizin zu bewerten.</p> <table border="0"> <tr> <td>Insgesamt: 1,45</td> <td>Standardabw.: 0,506</td> </tr> <tr> <td>Arzt: 1,5</td> <td>Standardabw.: 0,522</td> </tr> <tr> <td>Pflege: 1,29</td> <td>Standardabw.: 0,488</td> </tr> <tr> <td>MMI+Med: 1,63</td> <td>Standardabw.: 0,518</td> </tr> <tr> <td>MMI.: 1</td> <td>Standardabw.: 0</td> </tr> <tr> <td>Sonst: 1,5</td> <td>Standardabw.: 0,707</td> </tr> </table>	Insgesamt: 1,45	Standardabw.: 0,506	Arzt: 1,5	Standardabw.: 0,522	Pflege: 1,29	Standardabw.: 0,488	MMI+Med: 1,63	Standardabw.: 0,518	MMI.: 1	Standardabw.: 0	Sonst: 1,5	Standardabw.: 0,707
Insgesamt: 1,45	Standardabw.: 0,506												
Arzt: 1,5	Standardabw.: 0,522												
Pflege: 1,29	Standardabw.: 0,488												
MMI+Med: 1,63	Standardabw.: 0,518												
MMI.: 1	Standardabw.: 0												
Sonst: 1,5	Standardabw.: 0,707												

<p><b>Es war für mich einfach, das Tool zu verwenden.</b></p> <table border="1"> <caption>Data for: Es war für mich einfach, das Tool zu verwenden.</caption> <thead> <tr> <th>Gruppe</th> <th>Mittelwert</th> <th>Standardabweichung</th> </tr> </thead> <tbody> <tr> <td>MMI</td> <td>1,53</td> <td>0,671</td> </tr> <tr> <td>MMI+Med</td> <td>1,29</td> <td>0,488</td> </tr> <tr> <td>Pflege</td> <td>1,63</td> <td>0,916</td> </tr> <tr> <td>Arzt</td> <td>1,33</td> <td>0,577</td> </tr> <tr> <td>Insgesamt</td> <td>2</td> <td>0</td> </tr> </tbody> </table>	Gruppe	Mittelwert	Standardabweichung	MMI	1,53	0,671	MMI+Med	1,29	0,488	Pflege	1,63	0,916	Arzt	1,33	0,577	Insgesamt	2	0	<p>Es war für mich einfach, das Tool zu verwenden.</p> <table border="0"> <tr> <td>Insgesamt: 1,53</td> <td>Standardabw.: 0,671</td> </tr> <tr> <td>Arzt: 1,58</td> <td>Standardabw.: 0,669</td> </tr> <tr> <td>Pflege: 1,29</td> <td>Standardabw.: 0,488</td> </tr> <tr> <td>MMI+Med: 1,63</td> <td>Standardabw.: 0,916</td> </tr> <tr> <td>MMI.: 1,33</td> <td>Standardabw.: 0,577</td> </tr> <tr> <td>Sonst: 2</td> <td>Standardabw.: 0</td> </tr> </table>	Insgesamt: 1,53	Standardabw.: 0,671	Arzt: 1,58	Standardabw.: 0,669	Pflege: 1,29	Standardabw.: 0,488	MMI+Med: 1,63	Standardabw.: 0,916	MMI.: 1,33	Standardabw.: 0,577	Sonst: 2	Standardabw.: 0
Gruppe	Mittelwert	Standardabweichung																													
MMI	1,53	0,671																													
MMI+Med	1,29	0,488																													
Pflege	1,63	0,916																													
Arzt	1,33	0,577																													
Insgesamt	2	0																													
Insgesamt: 1,53	Standardabw.: 0,671																														
Arzt: 1,58	Standardabw.: 0,669																														
Pflege: 1,29	Standardabw.: 0,488																														
MMI+Med: 1,63	Standardabw.: 0,916																														
MMI.: 1,33	Standardabw.: 0,577																														
Sonst: 2	Standardabw.: 0																														
<p><b>Wie zufrieden sind Sie mit den gewählten Beschreibungen?</b></p> <table border="1"> <caption>Data for: Wie zufrieden sind Sie mit den gewählten Beschreibungen?</caption> <thead> <tr> <th>Gruppe</th> <th>Mittelwert</th> <th>Standardabweichung</th> </tr> </thead> <tbody> <tr> <td>MMI</td> <td>1,83</td> <td>0,922</td> </tr> <tr> <td>MMI+Med</td> <td>2</td> <td>0,775</td> </tr> <tr> <td>Pflege</td> <td>1,14</td> <td>0,378</td> </tr> <tr> <td>Arzt</td> <td>2,17</td> <td>1,282</td> </tr> <tr> <td>Insgesamt</td> <td>2</td> <td>1</td> </tr> </tbody> </table>	Gruppe	Mittelwert	Standardabweichung	MMI	1,83	0,922	MMI+Med	2	0,775	Pflege	1,14	0,378	Arzt	2,17	1,282	Insgesamt	2	1	<p>Wie zufrieden sind Sie mit den gewählten Beschreibungen?</p> <table border="0"> <tr> <td>Insgesamt: 1,83</td> <td>Standardabw.: 0,922</td> </tr> <tr> <td>Arzt: 2</td> <td>Standardabw.: 0,775</td> </tr> <tr> <td>Pflege: 1,14</td> <td>Standardabw.: 0,378</td> </tr> <tr> <td>MMI+Med: 2,17</td> <td>Standardabw.: 1,282</td> </tr> <tr> <td>MMI.: 2</td> <td>Standardabw.: 1</td> </tr> <tr> <td>Sonst: 2</td> <td>Standardabw.: 0</td> </tr> </table>	Insgesamt: 1,83	Standardabw.: 0,922	Arzt: 2	Standardabw.: 0,775	Pflege: 1,14	Standardabw.: 0,378	MMI+Med: 2,17	Standardabw.: 1,282	MMI.: 2	Standardabw.: 1	Sonst: 2	Standardabw.: 0
Gruppe	Mittelwert	Standardabweichung																													
MMI	1,83	0,922																													
MMI+Med	2	0,775																													
Pflege	1,14	0,378																													
Arzt	2,17	1,282																													
Insgesamt	2	1																													
Insgesamt: 1,83	Standardabw.: 0,922																														
Arzt: 2	Standardabw.: 0,775																														
Pflege: 1,14	Standardabw.: 0,378																														
MMI+Med: 2,17	Standardabw.: 1,282																														
MMI.: 2	Standardabw.: 1																														
Sonst: 2	Standardabw.: 0																														
<p><b>Die Anzahl der Bewertungsabstufungen...</b></p> <table border="1"> <caption>Data for: Die Anzahl der Bewertungsabstufungen...</caption> <thead> <tr> <th>Kategorie</th> <th>Anzahl Probanden</th> </tr> </thead> <tbody> <tr> <td>...wurde richtig gewählt</td> <td>29</td> </tr> <tr> <td>...ist zu hoch</td> <td>2</td> </tr> <tr> <td>...ist zu gering</td> <td>1</td> </tr> </tbody> </table>	Kategorie	Anzahl Probanden	...wurde richtig gewählt	29	...ist zu hoch	2	...ist zu gering	1	<p>Die Anzahl der Bewertungsabstufungen...</p> <table border="0"> <tr> <td>... wurde richtig gewählt:</td> <td>29</td> </tr> <tr> <td>... ist zu hoch:</td> <td>2 (Arzt, Arzt)</td> </tr> <tr> <td>... ist zu gering:</td> <td>1 (MMI)</td> </tr> </table>	... wurde richtig gewählt:	29	... ist zu hoch:	2 (Arzt, Arzt)	... ist zu gering:	1 (MMI)																
Kategorie	Anzahl Probanden																														
...wurde richtig gewählt	29																														
...ist zu hoch	2																														
...ist zu gering	1																														
... wurde richtig gewählt:	29																														
... ist zu hoch:	2 (Arzt, Arzt)																														
... ist zu gering:	1 (MMI)																														

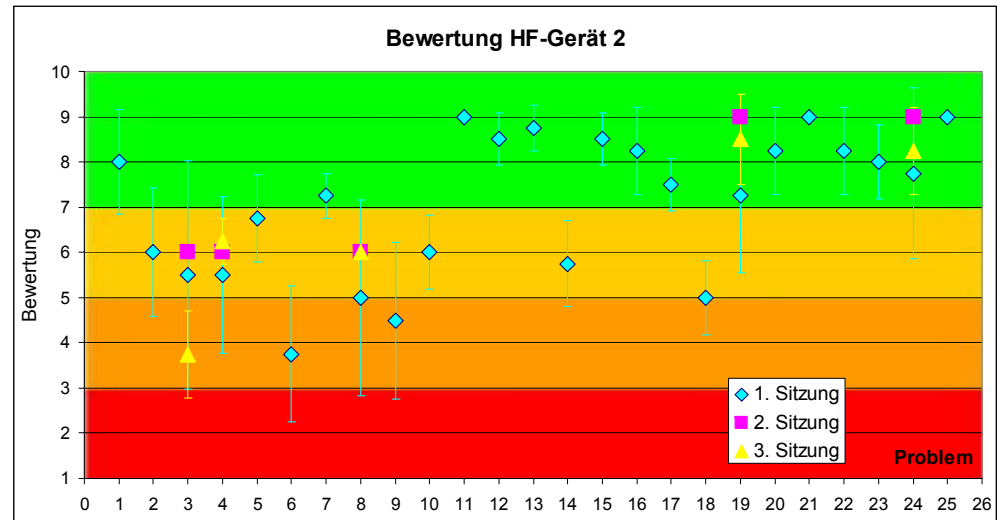
**Anhang N: Bewertung der HF Geräte 3. Evaluationsstudie**  
**1. HF-Gerät**

Problem	Bewertung			Gesamt- ergebnis	Reliabilitäts- workshop
	Sitzung 1	Sitzung 2	Sitzung 3		
1	7,25			7,25	6,75
2	4,75	3	6,25	6,25	6,5
3	6,75			6,75	7
4	3,75			3,75	5
5	6	7	5,5	5,5	5,33
6	6,5			6,5	7
7	3,5	7	6,75	6,75	7
8	3,5	6	6	6	4,5
9	7,5			7,5	8
10	7,5			7,5	7
11	8			8	6,75
12	8			8	8
13	9			9	9
14	8,5			8,5	7,75
15	8,25			8,25	6,75
16	8,75			8,75	8,5
17	6	9	9	9	8,75
18	9			9	8,75
19	7	6	7,25	7,25	6
20	7,25			7,25	7,5
21	6,5	9	8	8	6,75
22	8,75			8,75	7
23	8	7	7,25	7,25	7,5
24	8,75			8,75	9
25	7,25	8	8	8	7
26	8,5			8,50	8
27	7,75	9	7,5	7,50	7,25



## 2. HF-Gerät

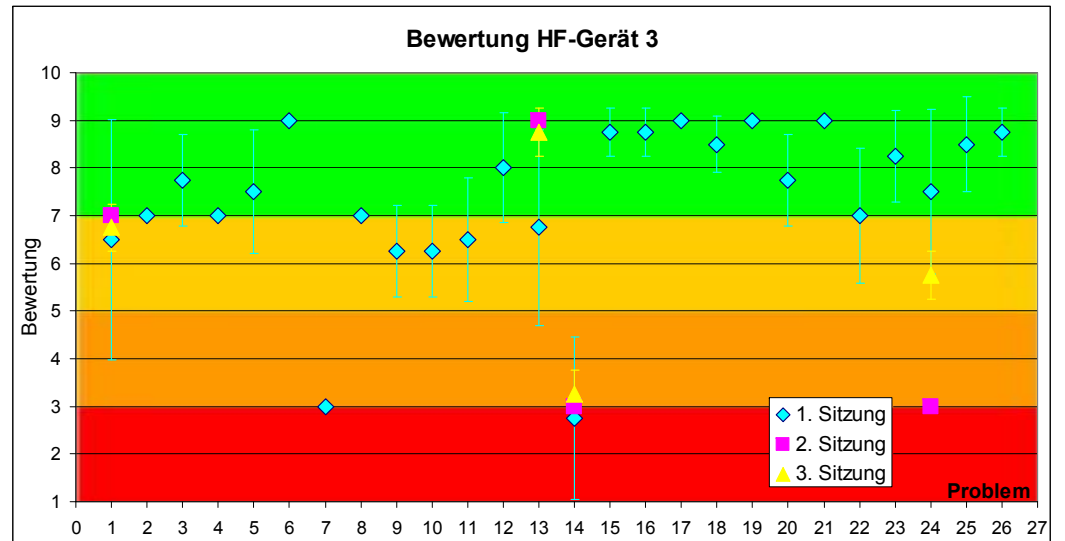
Problem	Bewertung			Gesamtergebnis	Reliabilitätsworkshop
	Sitzung 1	Sitzung 2	Sitzung 3		
1	8			8	
2	6			6	
3	5,5	6	3,75	3,75	4,75
4	5,5	6	6,25	6,25	6
5	6,75			6,75	
6	3,75			3,75	
7	7,25			7,25	
8	5	6	6	6	5,75
9	4,5			4,5	
10	6			6	
11	9			9	
12	8,5			8,5	
13	8,75			8,75	
14	5,75			5,75	
15	8,5			8,5	
16	8,25			8,25	
17	7,5			7,5	
18	5			5	
19	7,25	9	8,5	8,5	8,3125
20	8,25			8,25	
21	9			9	
22	8,25			8,25	
23	8			8	
24	7,75	9	8,25	8,25	8,3125
25	9			9	





### 3. HF-Gerät

Problem	Entwicklung			Gesamtergebnis	Reliabilitätsworkshop
	Sitzung 1	Sitzung 2	Sitzung 3		
1	6,5	7	6,75	6,5	6,6875
2	7			7	
3	7,75			7,75	
4	7			7	
5	7,5			7,5	
6	9			9	
7	3			3	
8	7			7	
9	6,25			6,25	
10	6,25			6,25	
11	6,5			6,5	
12	8			8	
13	6,75	9	8,75	6,75	7,8125
14	2,75	3	3,25	2,75	2,9375
15	8,75			8,75	
16	8,75			8,75	
17	9			9	
18	8,5			8,5	
19	9			9	
20	7,75			7,75	
21	9			9	
22	7			7	
23	8,25			8,25	
24	7,5	3	5,75	7,5	5,9375
25	8,5			8,5	
26	8,75			8,75	



## Anhang O: Gefundene Gebrauchstauglichkeitsprobleme und die dazugehörigen Arbeitsaufgaben (vgl. Kap. 6.2.4)

### Ständer bestücken (in Aufgabe 2 gefunden)

**Aufgabe:** Bestücken Sie den Ständer mit 3 Spritzenpumpen.

**Aufgabenerfüllung:** Bei sechs von neun Teilnehmern verkantete der Perfusor beim Einführen. Bei einer Testperson fiel das Gerät herunter und wurde beschädigt. Eine weitere Testperson drückte das Gerät gewaltsam in die Station, bis es einhakte. Obwohl eine der Führungsschienen nicht getroffen wurde, hakte das Gerät ein und funktionierte.

**Problem 1:** Der Perfusor kann beim Einführen verhaken (Vorkommnis: 67 %).

### Einschalten (in Aufgabe 3)

**Aufgabe:** Schalten Sie die oberste Spritzenpumpe im Ständer ein.

**Aufgabenerfüllung:** Zwei Probanden drückten zuerst „Start“ anstelle des Power-Buttons. Dies führte zu einer Verzögerung von ca. 10 Sekunden.

**Problem 2:** Der Unterschied zwischen Start und Power ist nicht klar (Vorkommnis: 22%)

### Spritze einlegen/Spritzenwechsel (in Aufgabe 4 gefunden)

**Aufgabe:** Bestücken Sie die betriebsbereite Spritzenpumpe mit „Spritze 1“. Folgen Sie dazu den Anweisungen der Spritzenpumpe zur Spritzenauswahl (3x gestellt).

**Aufgabenerfüllung:** Fast jede Testperson hatte Probleme beim Einlegen der Spritze. Die Bedienung des Spritzenbügels sowie die vorzunehmende Lage/Positionierung der Spritze waren unklar. Nach dem Einlegen fragt das Gerät nach Sprizentyp und -größe. Da die Spritze hinter dem Bedienpanel nicht mehr einsehbar ist, mussten viele Testnutzer sie erneut entnehmen.

**I.) Problem 3:** Spritze bei vorgefahrenem Antrieb eingelegt, Bügel nicht umgelegt und lange im Menü gesucht.

**Aufgabenerfüllung:** Es war nicht ersichtlich, dass der Bügel zum Spritzenwechsel umgelegt werden musste (Vorkommnis: 44%)

**II.) Problem 4:** Probleme beim Einlegen der Spritze

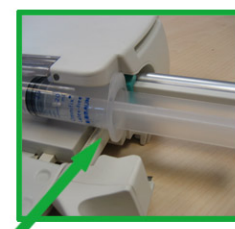
**Aufgabenerfüllung:** Probanden hatten Probleme, die Spritze einzulegen. Sie trafen die Spritzenfixierung nicht, bzw. ihnen war diese Funktion nicht ersichtlich (Vorkommnis: 78%).

**III.) Problem 5:** Leitung eingeklemmt

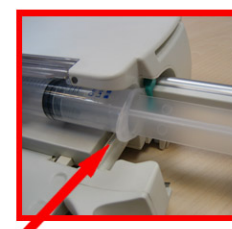
**Aufgabenerfüllung:** Viele der Probanden klemmten die Leitung beim Einlegen der Spritze ein. In einigen Fällen konnte hierdurch die Spritzenklappe nicht geschlossen werden. Wenn die Leitung so eingeklemmt wird, dass dies Einfluss auf die Förderung hat, erscheint eine Hinweismeldung „Druck zu hoch“. (Vorkommnis: 56%).

**IV.) Problem 6:** Spritze horizontal eingelegt

**Aufgabenerfüllung:** Es ist möglich, eine Spritze auch horizontal einzulegen. Wenn z.B. eine Omnifix 20ml Spritze horizontal eingelegt wird, kann in bestimmten Positionen im Display „Omnifix 30ml“ erscheinen. Wenn diese Option ausgewählt wird, erhält der Patient 25% weniger an Medikamentendosis. Ein Proband versuchte die Spritze horizontal einzulegen; er bemerkte diesen Irrtum jedoch selbst bei der Spritzenauswahl (Vorkommnis: In der Expertenevaluation und ein Proband legte die Spritze falsch ein und bemerkte den Fehler, da keine Omnifix-Spritze angezeigt wurde).



richtig



falsch

**Rate eingeben und Förderung starten (in Aufgabe 5 und 6 gefunden)****Aufgaben:**

1. Geben Sie eine Rate von 10 ml/h ein.
2. Geben Sie die Dosierung für das Medikament anhand folgender Variablen ein:  
Volumen = 30ml, Dauer = 4 Stunden und starten Sie die Förderung.

**I.) Problem 7: „OK-Button“ statt „Start-Button“ betätigt**

**Aufgabenerfüllung:** Probanden bestätigten Werte mit „OK“. Sie dachten, dass sie mit diesem Button auch starten können und bemerkten erst später, dass hierfür das Betätigen des „Start-Buttons“ erforderlich ist (Vorkommnis: 56%).

**II.) Problem 8: Einzelschritte zur Werteeingabe**

**Aufgabenerfüllung:** Um Werte einzustellen, verwendeten die Probanden Einerschritte. Sie erkannten nicht, dass auch die Zehnerstelle ausgewählt werden kann (Vorkommnis: 56%).

**Bolusgabe (in Aufgabe 7 und 14)****Aufgaben:**

- 1.) Die Schmerzen des Patienten verstärken sich plötzlich. Verabreichen Sie einen Bolus von 6 ml.
- 2.) Die Schmerzen des Patienten lassen auch nach erhöhter Dosierung nicht nach. Leiten Sie eine Bolusgabe von 7 ml in 14 Sekunden ein und verabreichen Sie diese. Danach fahren Sie mit der vorherigen Dosierung fort.

**I.) Problem 9: START statt „BOL“ drücken um Bolus zu verabreichen.**

**Aufgabenerfüllung:** Probanden wählten richtig mit „BOL“ den Bolus aus, stellten Werte ein und drückten START/STOP (wie bei normaler Förderung), um den Bolus zu verabreichen. Es müsste jedoch wiederum BOL gedrückt werden. Bei START/STOP blieb nicht nur der Bolus aus, sondern die gesamte Förderung wird abgebrochen (Vorkommnis: 22%).

**II.) Problem 10: „OK“ gedrückt um Bolus zu starten.**

**Aufgabenerfüllung:** Probanden wählten richtig mit „BOL“ den Bolus aus, stellten Werte ein und drückten dann „OK“ um den Bolus zu verabreichen. Nach der Bol-Dosis gelangt man zur Bol-Zeit. Wenn dann „OK“ gedrückt wird, erscheint ein Hinweis „BOL drücken zum Start des Bolus“. Solange dieser Hinweis erscheint (2-3 Sek.), ist keine Taste aktiv (Vorkommnis: 56%).

**III.) Problem 11: Im Auswahlmeneü „BOL“ gedrückt.**

**Aufgabenerfüllung:** Ein Proband gelangte durch das vorherige Problem in das Hauptmenü, drückte dann erneut Bolus („BOL“), um diesen zu verabreichen und gelangte hierdurch in ein Menü „Leitung entlüften“. Er dachte, er müsste die Leitung entlüften. Entlüftete die Leitung trotz eines Hinweises „Vor dem Entlüften Patienten von der Pumpe trennen“. Dies geschah 11mal mit je 1ml (Vorkommnis: 11%).

**IV.) Problem 12: Kleiner Pfeil zur Rateneinstellung**

**Aufgabenerfüllung:** Von einigen Probanden wurde zuerst nicht erkannt, wie die Rateneinstellung auszuwählen ist. Der kleine Pfeil, der einen Hinweis auf das Drücken der „Links-Taste“ gibt, wurde nicht erkannt. (Vorkommnis: 33%).

**Kontrasteinstellung und Standard-Bolusgabe (in Aufgabe 10 und 16 gefunden)****Aufgabe:**

- 1.) Stellen Sie den Kontrast und die Display-/Tastaturbeleuchtung so ein, dass es für Sie angenehm ist.
- 2.) Ändern Sie die Voreinstellung der Standard-Bolusgabe auf 600 ml/h.

**Problem 13:** Aufruf des Hauptmenüs durch [c] und weitere Menüs über die Scrollfunktion. Unübersichtlich und schwer zu erkennen (Vorkommnis: 44 %).

**Aufgabenerfüllung:** Einige Testnutzer wussten nicht, wie man das Hauptmenü aufruft. Anstatt bei laufender Förderung die C-Taste zu drücken, stoppte ein Testnutzer die Förderung, um die gewünschten Einstellungen vornehmen zu können. Als Problem wurde auch hier oft die Orientierung im Menü genannt; Testnutzer wussten nicht, wo sie sich befanden. Das Hauptmenü wurde nicht als solches erkannt. Die Testnutzer erkannten hier auch nicht die Scrollfunktion.

**Fehlendes Feedback (Globales Problem)**

**Problem 14:** Einige Probanden bemängelten ein fehlendes Feedback, z.B.:

- bei der Übernahme von Werten oder der Rateneinstellung und bei der Bolusgabe

## Anhang P: Problembewertung bei den entsprechenden Arbeitsaufgaben (vgl. Kap. 6.2.4)

Aufgabe	Problem	Problembeschreibung	Bewertungsnote
1. Aufgabe	<b>Spacestation bestücken</b>	1 Perfusor verkantet, Funktionsschiene verfehlt	3
2. Aufgabe	<b>Einschalten</b>	2 Start gedrückt statt Power	8,4
3. Aufgabe	<b>Spritze einlegen / wechseln</b>	3 Spritze bei vorgefahrenem Antrieb eingelegt/Bügel nicht umgelegt	4,8
		4 Naht nicht getroffen / Probleme beim Einlegen	6
		5 Leitung eingeklemmt	6
		6 Spritze horizontal eingelegt	2,4
4. Aufgabe	<b>Rate eingeben und Förderung starten</b>	7 „OK-Button“ statt „Start-Button“ betätigt	6,8
		8 Einzelschritte zur Werteeingabe	7,2
5. Aufgabe	<b>Bolusgabe</b>	9 START gedrückt um Bolus zu verabreichen	1,4
		10 OK gedrückt um Bolus zu starten	3
		11 Im Auswahlmenü Bolus gedrückt	2
		12 Kleiner Pfeil zur Rateneinstellung	8,4
6. Aufgabe	Sonderfunktionen: Kontrasteinstellung u. Standard-Bolusgabe	13 Im Hauptmenü ist die „Scrollfunktion“ nicht klar	6,2
Global	Fehlendes Feedback	14 Fehlendes Feedback	1,6

Die Note für das Problem Nr. 9 „Start gedrückt, um Bolus zu verabreichen“ bestimmt die Gebrauchstauglichkeitsbewertung für das gesamte System, da es sich hierbei um ein sicherheitskritisches Problem handelt.

## Anhang Q: Testaufgaben Herzkatheterpumpe (vgl. Kap. 6.4)

### Aufgabenbeschreibung Test 1:

#### Vorbereitung

Auf dem Flur: „Frau Podolski ist 56 Jahre alt und ist gerade mit einem kardiogenen Schock eingeliefert worden. Ihre Kollegen und Sie haben sich dazu entschieden, Frau Podolski mit der RCP zu behandeln. Bitte überprüfen Sie, ob die hierfür notwendigen Materialien bereitliegen. Fragen Sie hierfür gegebenenfalls Ihren Assistenten. In der folgenden Behandlung werden Sie die Rolle des unsterilen Gerätebedieners übernehmen. Alle Aufgaben an der Konsole führen Sie selbstständig durch. Der Assistent wird Ihnen hierbei nicht helfen. Er wird lediglich die Aufgaben im sterilen Bereich übernehmen und Sie ggf. über Vitalparameter des Patienten informieren.“

#### Szenario 1: Setup durchführen

Testleiter kommt in den Flur gelaufen. „Der Patient ist vorbereitet und narkotisiert. Blutdruck ist schwach – 85 zu 60. Puls 110. Wir müssen schnell die Pumpe anschließen.“

#### Szenario 2: RCP starten (8000 U/min)

Off-Sprecher: „Stellen Sie die Pumpe auf 8000 U/min“.

#### Szenario 3: Helligkeit erhöhen

Assistent: „Ich kann von hier aus die Werte nicht genau erkennen. Stellen Sie das bitte einen Schritt heller“.

→ ASQ für Szenario 1, 2 und 3 ausfüllen

#### Szenario 4: Rotorgeschwindigkeit erhöhen (10.000)

Assistent: „Die Vitalparameter haben sich verbessert. Blutdruck ist aber immer noch schwach. Wir sollten auf 10.000 Umdrehungen erhöhen.“

→ Während der Einstellung legt der sterile Assistent ein OP-Tuch über den Motor.

#### Szenario 5: Warnmeldung Temperatur

→ Warnmeldung über zu hohe Motortemperatur erhalten

#### Szenario 6: Alarm history

Assistent: „Da waren schon einmal ein paar Warnmeldungen. Können Sie mal herausfinden, was da los war?“

#### Szenario 7: Notaus

„Da stimmt etwas nicht, machen Sie schnell mal das Ding aus.“

→ ASQ für Szenario 4 bis 7 ausfüllen

#### Szenario 8: RCP starten

Assistent: „Es ist alles wieder OK. Wir können mit der Behandlung weitermachen. Also 10.000 Umdrehungen pro Minute.“

#### Szenario 9: Tastensperre aktivieren

Assistent: „So ich denke jetzt ist alles in Ordnung. Stellen Sie mal die Tastensperre ein, dass da keiner ungewollt etwas verstellt.“

#### Szenario 10: Spüllösung erhöhen (1.300 ml/24h)

Assistent: „Ich glaube, wir sollten doch noch die Spüllösung etwas hochfahren. Stellen Sie die mal auf 1.300 ml.“

#### Szenario 11: Hilfe aufrufen

Assistent: „Wie ist das jetzt eigentlich, wenn ich morgen mit jemand anderem arbeite? Kann ich da auch so etwas wie eine Hilfe aufrufen?“

#### Szenario 12: RCP stoppen und Konsole herunterfahren

Assistent: „So, ich glaube, das war alles. Der Blutdruck ist stabil. Bitte fahren Sie die RCP runter. Stellen Sie die WeanOff-Zeit auf eine Minute.“

→ ASQ für Szenario 8 bis 12 ausfüllen & Fragebögen

## Aufgabenbeschreibung Test 2

Hier liegt unser Patient Herr Krüger, 56 Jahre. Er wurde vor einer Stunde mit einem kardiogenen Schock eingeliefert. Wir werden ihm jetzt eine Reitan Katheterpumpe implantieren. Sie arbeiten im unsterilen Bereich und bedienen dort die Konsole. Ich übernehme die Arbeit im sterilen Bereich am Patienten.

Ich bin ebenso wie Sie noch unerfahren im Umgang mit dem System und möchte wissen, wie Sie, als unerfahrener Nutzer, mit dem System zurechtkommen.

Bitte bedenken Sie, dass nicht Sie getestet werden sondern das Gerät. Fehler, die im Testverlauf auftreten, liefern uns wertvolle Informationen zur Verbesserung des Geräts.

Darum ist es auch wichtig, dass Sie uns alle Ihre Überlegungen und Gefühle laut mitteilen. Bitte kommentieren Sie Ihre Handlungen während des Tests laut und drücken Sie aus, worüber Sie sich ärgern oder freuen. So können wir besser verstehen, was an dem Gerät noch zu verbessern ist oder was schon gut gelungen ist.

Sie sehen hier einen Fragebogen zu den einzelnen Aufgaben. Wir werden Sie in unregelmäßigen Abständen bitten, diesen auszufüllen. Während des Ausfüllens können Sie gerne etwas trinken oder essen.

Wenn Sie jetzt keine Fragen mehr haben und Ihnen alles klar ist, beginnen wir nun den Test mit der ersten Aufgabe.

### **Aufgabe 1: Installation**

Zu Beginn müssen wir das System installieren.

Versuchen Sie bitte jetzt, mit Hilfe der Anweisungen, die Ihnen das System gibt, die Herzkatheterpumpe zu installieren.

→ ASQ für Szenario 1 ausfüllen

### **Aufgabe 2: Inbetriebnahme/Propellergeschwindigkeit**

Nun nehmen Sie bitte die Katheterpumpe in Betrieb. Wie schnell dreht sich denn der Propeller gerade?

Dann erhöhen Sie bitte die Geschwindigkeit auf 5.000 [U/min].

### **Aufgabe 3: Helligkeit**

Ich kann die Angaben auf dem Bildschirm nicht richtig erkennen. Bitte schauen Sie doch mal, ob Sie die Helligkeit des Bildschirms anpassen können.

→ Motor wird mit Bauchtuch abgedeckt.

### **Aufgabe 4: Warnmeldung Motor**

Bitte schauen Sie doch mal, ob man die Angaben auf dem Bildschirm auch aus größerer Entfernung noch erkennen kann.

→ Warnmeldung für Temperatur erscheint und muss behoben werden

→ ASQ für Szenario 2, 3 und 4 ausfüllen

### **Aufgabe 5: Spüllösung**

Wie hoch ist denn die Menge der Spüllösung? Bitte erhöhen Sie sie um 200 [ml/24h].

### **Aufgabe 6: Notaus**

Ich glaube, hier gibt es ein Problem. Die Herzkatheterpumpe scheint sich verkantet zu haben. Oh nein, schnell, machen Sie den Propeller aus!

### **Aufgabe 7: Wiederinbetriebnahme der Pumpe**

Ich glaube, ich habe mich geirrt. Sie können die Herzkatheterpumpe wieder in Betrieb nehmen. Bitte erhöhen Sie die Geschwindigkeit des Propellers wieder auf 5000 [U/min].

→ ASQ für Szenario 5, 6 und 7 ausfüllen

**Aufgabe 8: Technischer Status**

Nach der Warnmeldung vorhin würde ich gerne die Motortemperatur überprüfen. Bitte sagen Sie mir, wie hoch die Motortemperatur ist.

**Aufgabe 9: Warnmeldung Propellergeschwindigkeit**

Herrn Krügers Werte haben sich deutlich verbessert. Bitte verringern Sie darum die Propellergeschwindigkeit auf 2000 [U/min].

→ Warnmeldung für zu geringe Rotorgeschwindigkeit erscheint.  
Problem muss behoben werden

(*Testperson fragt, was Sie tun soll.* → Dann reagieren Sie bitte entsprechend.)

**Aufgabe 10: Alarmverlauf**

Das war bereits die zweite Warnmeldung. Mich interessiert, ob die Alarmmeldungen noch irgendwo gespeichert sind. Bitte schauen Sie mal nach, ob wir diese abrufen können.

→ ASQ für Szenario 8, 9 und 10 ausfüllen

**Aufgabe 11: Hilfe**

Sie kennen sich ja schon ganz gut aus.

Aber wenn ich morgen mit einem anderen Kollegen zusammenarbeiten muss, der sich nicht so gut auskennt, würde ich gerne wissen, ob das System bei Problemen Hilfe gibt.

Bitte suchen Sie doch mal, ob es diese Möglichkeit gibt.

**Aufgabe 12: Stopp und Weaning**

Der Allgemeinzustand des Patienten hat sich gebessert. Bitte beenden Sie den Einsatz der Pumpe mit Weaning. Lassen Sie den Propeller bitte eine Minute auslaufen.

**Aufgabe 13: Exit**

Gut, jetzt können Sie das System beenden.

→ ASQ für Szenario 11, 12 und 13 ausfüllen



## Anhang R: Herzkatheterpumpe – Probleme beider Studien (vgl. Kap. 6.4)

Problem	Problem gefunden in...	
	Studie 1	Studie 2
1	Levensensor am Gerät gesucht	X X
2	Nicht den Klebestreifen am Levensensorpad erkannt	X X
3	Levensensorpad nicht gefunden	X
4	Nicht klar, wo Levensensor festgemacht wird	X X
5	Levensensor nicht angeschlossen, da Schritt übersprungen	X
6	Levensensor schwer einzustecken	X X
7	Hält Drive Unit die ganze Zeit in der Hand	X X
8	Übergabe Schlauchleitung. Unklarheit, was mit Schlauchleitung gemeint ist	X
9	Leitungen trennen und auf Patientenlevel halten zuerst nicht ganz klar	X X
10	Tropfenzähler und Rückführbeutel werden in der Hand gehalten bis Arbeitsanweisung erscheint.	X X
11	Entlüften dauert lange	X X
12	Entlüften: es wird nicht verstanden, dass das automatisch funktioniert	X
13	Umständliches Bücken zum Anschließen der Leitungen unter dem Gerät	X X
14	Stromkabel beim Anschließen des Drive Unit berührt --> rutscht heraus --> Stromversorgung getrennt	X
15	Bezeichnung Drive Unit und RCP nicht ganz klar	X X
16	Rückführbeutel nicht erkannt	X X
17	Katheter und Schlauchset verwechselt	X X
18	„Ja“ und „Nein“ verwechselt bei Abfrage, ob Katheter reibungslos läuft	X
19	Leitungen trennen übersprungen	X
20	Leitungen selbst getrennt und daher keine sterile Übergabe	X X
21	Kassettschritt übersprungen, daher nicht sofort gesteckt	X
22	Probleme, die Leitungen an die RCP anzuschließen	X
23	Erkennt linke Seite nicht, drückt RCP-Screen	X X
24	Pfeiltasten zu sensibel	X X
25	Es ist unklar, was der Motor ist.	X
26	Probanden drückten Warnmeldung frühzeitig weg.	X
27	Rotorleistung wird verdeckt.	X
28	Proband will Propeller auf 3000 laufen lassen, obwohl Warnmeldung noch über dem Fenster eingeblendet ist.	X
29	Erst auf Exit gedrückt	X
30	Unklar, ob Temperaturangabe sich auf die Motortemperatur bezieht	X
31	Alarmverlauf nicht gefunden	X
32	Tastensperre nicht gefunden	X
33	Fragezeichen nicht gesehen	X X
34	Spüllösung einstellen. Es wird im Technischen Verlauf nachgesehen. Dort keine Möglichkeit	X
35	Wusste nicht, was sich hinter dem Fragezeichen verbirgt	X
36	Durch Pfeiltasten Konsole stoppen	X X
37	WeanOff läuft schon, wenn das WeanOff-Menü geöffnet wird	X X
38	Zeiteinstellung nicht klar	X X
39	Spüllösung gestoppt, bevor Katheter entfernt wurde	X X
40	Katheter entfernt, bevor der Käfig geschlossen wurde	X X
41	„Stop“ ist nicht gleich WeanOff	X X
42	Ärger über Tastensperre. Wunsch, diese ausschalten zu können.	X X

Markierungen für die 15 und 20 wichtigsten Probleme ausgewählt durch den Testleiter

Problem 15 wurde in Studie 2 als zwei getrennte Probleme angesehen und beide für Top 15 ausgewählt

Aufgaben 27, 28 und 30 wurden nicht in beiden Studien durchgeführt.

## Anhang S: Alle Bewertungsergebnisse für die Katheterpumpe

Bewertung Studie 1				Bewertung Studie 2			
Problem	Beschreibung	Mittelwert	Sta. Abw.	Mittelwert	Sta. Abw.	Problem	Beschreibung
2	Nicht den Klebestreifen am Levelsensorpad erkannt	7,6	0,548	7,6	0,548	1	Klebestreifen nicht erkannt
3	Nicht klar, wo Levelsensor festgemacht wird.	7,2	1,095	7	0,000	2	Pad nicht unten angeklebt
6	Hält Drive Unit die ganze Zeit in der Hand	7,8	0,447	6,8	0,447	7	Drive Unit in der Hand
7	Leitungen trennen und auf Patientenlevel halten nicht ganz klar	9	0,000				
9	Umständliches Bücken zum Anschließen der Leitungen	6,8	0,447	6,2	0,837	3	Anschlüsse unten - Bücken
12	Rückführbeutel nicht erkannt	7,6	0,548	7,8	0,837	6	Rückführbeutel nicht erkannt
14	„Ja“ und „Nein“ verwechselt bei Abfrage, ob Katheter reibungslos läuft	7,2	0,447				
20	Erkennt linke Seite nicht, drückt RCP-Screen	3,6	1,342	5,2	0,447	9	Inbetriebnahme (Start)
22	Es ist unklar, was der Motor ist	6,8	0,447				
24	Rotorleistung wird verdeckt.	8,4	0,894				
25	Erst auf Exit gedrückt.	1,4	0,548	1,4	0,548	10	Drücken auf Exit für Notaus
26	Fragezeichen nicht gesehen	6,2	0,837	7,2	1,095	11	Hilfe gesucht
29	WeanOff läuft schon, wenn das WeanOff Menü geöffnet wird	6	0,707	7	0,000	13	Weanoff läuft schon, wenn WeanOff geöffnet
30	Zeiteinstellung nicht klar	5,8	0,447	6,6	0,548	14	Zeiteinstellung nicht klar
31	Spüllösung gestoppt, bevor Katheter entfernt wurde	3,8	1,643	9	0,000	16	Spülfluss beendet, bevor Katheter entfernt
32	Katheter entfernt, bevor der Käfig geschlossen wurde	2	0,000	2	0,000	15	Katheter entfernt, bevor Käfig geschlossen
33	„Stopp“ ist nicht gleich WeanOff	7	0,000	6,2	0,447	12	„Stopp“ ist nicht gleich „Weanoff“
				6	0,000	4	Netzstecker nicht gesichert
				6,2	0,447	5	Konnektieren der Anschlüsse
				6,4	0,894	8	Leitungen trennen

Fettdruck besagt: von beiden Studien bewertet und für die Korrelationsmessung herangezogen

## Anhang T: Vergleich Studie „Katheterpumpe“

Problem	Mittelwert der Expertenbewertung		Mittelwert	Differenz ( $\bar{d}$ )
	Studie 1	Studie 2		
1	7,6	7,6	7,6	0
2	7,2	7	7,1	0,2
3	7,8	6,8	7,3	1
4	6,8	6,2	6,5	0,6
5	7,6	7,8	7,7	-0,2
6	3,6	5,2	4,4	-1,6
7	1,4	1,4	1,4	0
8	6,2	7,2	6,7	-1
9	6	7	6,5	-1
10	5,8	6,6	6,2	-0,8
11	3,8	9	6,4	-5,2
12	2	2	2	0
13	7	6,2	6,6	0,8
Mittelwert der Messwertdifferenzen ( $\bar{d}$ )				-0,5538
Standardabweichung				1,5899

Korrelationen		Studie1	Studie2
Studie1	Korrelation nach Pearson	1	,734**
	Signifikanz (2-seitig)		,004
	N	13	13

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

## Anhang U – Bewertungsergebnisse zur Reliabilitätsberechnung und Bland-Altman-Methode

Problem	Studie 1	Studie 2	Differenz	Mittelwert
1	7,25	6,75	0,5	7
2	6,25	6,5	-0,25	6,375
3	6,75	7	-0,25	6,875
4	3,75	5	-1,25	4,375
5	5,5	5,33333333	0,1667	5,4167
6	6,5	7	-0,5	6,75
7	6,75	7	-0,25	6,875
8	6	4,5	1,5	5,25
9	7,5	8	-0,5	7,75
10	7,5	7	0,5	7,25
11	8	6,75	1,25	7,375
12	8	8	0	8
13	9	9	0	9
14	8,5	7,75	0,75	8,125
15	8,25	6,75	1,5	7,5
16	8,75	8,5	0,25	8,625
17	9	8,75	0,25	8,875
18	9	8,75	0,25	8,875
19	7,25	6	1,25	6,625
20	7,25	7,5	-0,25	7,375
21	8	6,75	1,25	7,375
22	8,75	7	1,75	7,875
23	7,25	7,5	-0,25	7,375
24	8,75	9	-0,25	8,875
25	8	7	1	7,5
26	8,5	8	0,5	8,25
27	7,5	7,25	0,25	7,375
	Mittelwert		0,33950617	
	Standardabweichung		0,73812751	

### Korrelationen

		BewertungS1	BewertungS2
BewertungS1	Korrelation nach Pearson	1	,809**
	Signifikanz (2-seitig)		,000
	N	27	27
BewertungS2	Korrelation nach Pearson	,809**	1
	Signifikanz (2-seitig)	,000	
	N	27	27

\*\* Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

## Anhang V – Gesamtbewertung der „Katheterpumpe“

Aufgabe	Bewertung Studie 1			Bewertung Studie 2		
	Note des schlimmsten Problems	Bewertungsfaktor	Summe	Note des schlimmsten Problems	Bewertungsfaktor	Summe
<b>Setup</b>	6,8	4	27,2	6,2	4	24,8
<b>RCP starten</b>	3,6	4	14,4	5,2	4	20,8
Spüllösung verändern	9	1	9	9	1	9
<b>Warnmeldung Temperatur</b>	9	4	36	9	4	36
<b>Warnmeldung Geschwindigkeit</b>				9	4	36
<b>Geschwindigkeit verändern</b>	9	4	36	9	4	36
Technischer Status				9	1	9
Verlauf						
Alarm-Verlauf	9	1	9	9	1	9
Helligkeit	9	1	9	9	1	9
Kontrast						
Lautstärke	9	1	9	9	1	9
Service						
Hilfe	6,2	1	6,2	7,2	1	7,2
Tastensperre	9	1	9			
<b>Weaning</b>	5,8	4	23,2	6,2	4	24,8
<b>Notaus</b>	1,4			1,4		
<b>Ausschalten</b>	2			2		
		26	188		30	230,6
<b>Bewertung</b>	1,4		1,4	1,4		1,4
Bewertung ohne schwerwiegende Probleme	7,76		7,23	8,07		7,68

Es werden nur die Aufgaben bewertet, die auch durchgeführt wurden. Grün bedeutet: kein Problem aufgefallen, daher 9

## Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Angestellter am Experimental-OP des Universitätsklinikums Tübingen. Im Rahmen dieser Arbeit habe ich von vielen Seiten Hilfestellung bekommen.

Meinem Doktorvater, Herrn Prof. Dr. Claussen, danke ich für die wissenschaftliche Unterstützung durch anregende Diskussionen und konstruktive Ratschläge, die es mir ermöglichten, diese interdisziplinäre Arbeit zu verfassen.

Meinem Betreuer, Herrn Dr. Matern, danke ich besonders für die intensive Zuwendung und die Möglichkeit, den Experimental-OP als Forschungsplattform zu benutzen. Die vielen Diskussionen, ob im Büro, auf langen Fahrten oder Kongressen, in denen wir nicht selten unterschiedlicher Meinung waren, haben sehr zum Gelingen dieser Arbeit beigetragen. Für die differenzierten Blickwinkel, die fachliche Betreuung sowie die kritische Durchsicht meiner Arbeit bin ich Herrn Dr. Matern sehr dankbar.

Weiterhin bedanke ich mich bei meinen Kollegen des Experimental-OPs, auf deren Rat und Wissen ich während der gesamten Arbeit zurückgreifen konnte. Dieses Wissen hat mir bei dem interdisziplinären Thema sehr geholfen. Hervorheben möchte ich hier Herrn Martin Scherrer für die Einblicke in die Krankenhaushygiene, Gebäude- und Medizintechnik, Frau Diana Hagen für die Tipps aus erster Hand im Bereich des medizinischen Arbeitsfeldes und Herrn David Voss für Informationen aus dem Bereich der medizinischen Informatik.

Weiterhin haben mich auf meinem Weg zum Abschluss dieser Arbeit viele interessante Menschen begleitet, die mich inspiriert, angeregt, unterstützt oder ermutigt haben. Besonders hervorheben möchte ich hierbei die aufschlussreichen Diskussionen mit Frau Prof. Dr. Susan Hallbeck, Herrn Prof. Dr. Jürgen Held, Frau Prof. Dr. Heidi Krömker, Herrn Prof. Gottfried Zimmermann, Frau Dr. Sibylle Hildenbrand, Herrn Dr. Klaus Neuder und Frau Dr. Marilyn Sue Bogner, die mir im Rahmen meiner Arbeit weitergeholfen haben. Ferner möchte ich mich bei den Studenten Thomas Baumann, Arvid Braumann, Phillipp Gonsor, Bernd Hallabrin, Anja Hültner, Christoph Kaminski, Felix Kleinert, Julia Kranz und Eva Schiendzielorz für Ihre Unterstützung bedanken. Teilergebnisse Ihrer Arbeiten sind in dieser Dissertation berücksichtigt.

Herrn Dr. Christoph Meisner danke ich für die kompetente statistische Beratung.

Weiterer Dank gilt auch allen Experten und Probanden, die im Verlauf dieser Arbeit an Befragungen, Gebrauchstauglichkeitstests, Expertenworkshops oder Bewertungen teilgenommen haben. Ohne ihre hilfsbereite Unterstützung hätte diese Arbeit nicht zufriedenstellend beendet werden können. Viele von ihnen haben zum größten Teil unentgeltlich ihre Freizeit geopfert, um mich im Rahmen dieses Themas zu unterstützen.

Ein besonderer Dank gilt meiner Mutter Hedi sowie meiner Freundin Eva Weichler, die mir immer zur Seite standen und einen wesentlichen Teil zum Gelingen dieser Arbeit beitrugen.

Der größte Dank gilt meinem Vater Horst Büchel, der mich während der ganzen Arbeit selbstlos unterstützt hat und viel Zeit und Mühen in diese Arbeit investiert hat.