

Statistische Analyseverfahren

Michael Hanslmaier und Dirk Baier

1 Einleitung

Bei der Analyse von Daten aus Opferbefragungen ist man oft mit der Tatsache konfrontiert, dass die lineare Regression mittels der Methode der kleinsten Quadrate (*Ordinary Least Squares* – OLS) aufgrund der Beschaffenheit der Daten nicht angewendet werden kann. Dies ist der Fall, wenn die abhängige Variable nicht metrisch, sondern nur nominalskaliert ist, wie etwa Prävalenzen, also die Angabe, ob eine Person in einem bestimmten Zeitraum Opfer geworden ist oder nicht. Auch kann die abhängige Variable sehr asymmetrisch, d. h. schief verteilt sein. Dies trifft beispielweise auf Inzidenzen zu, also die Häufigkeit der Opferwerdung. Darüber hinaus weisen Daten aus Opferbefragungen häufig eine hierarchische Struktur auf, d. h., die Befragten sind Teil übergeordneter Kontexte, etwa Schülerinnen und Schüler in Schulklassen oder Bewohnerinnen und Bewohner in Stadtvierteln. Diese auch als Cluster bezeichnete Datenstruktur kann die Folge eines mehrstufigen Stichprobenverfahrens sein, bei dem z. B. zunächst Stadtviertel ausgewählt werden und innerhalb dieser Stadtviertel dann Personen. Darüber hinaus kann es auch bei einfachen Zufallsstichproben von Interesse sein, Kontexte zu berücksichtigen, wenn diese die abhängige Variable beeinflussen (z. B. Eigenschaften von Stadtvierteln auf das Opferrisiko). Die Elemente eines Kontexts sind sich in der Regel ähnlicher als Elemente aus verschiedenen Kontexten, u. a. da diese den gleichen Umwelteinflüssen unterliegen. Infolgedessen sind die Beobachtungen nicht mehr statistisch unabhängig voneinander – eine Basisannahme der meisten statistischen Verfahren (Rabe-Hesketh/Skrondal 2012, 1–2; Windzio 2008, 113–114; ausführlicher zur Thematik komplexer Stichproben Lee/Forthofer 2006).

Der vorliegende Beitrag verfolgt zwei Ziele. Zum einen sollen statistische Verfahren für die multivariate Analyse von binären Daten (Prävalenzen), Zähldaten (Inzidenzen) und Mehrebenen Daten (Personen in Kontexten) vorgestellt werden. Der Vorteil multivariater Analysen liegt in der Möglichkeit, Drittvariablen kontrollieren und den jeweiligen Effekt verschiedener unabhängiger Variablen simultan schätzen zu können. Die Darstellung soll den Leserinnen und Lesern ein grundlegendes Verständnis der Verfahren liefern. Zum anderen sollen die Verfahren anhand von Beispielanalysen anschaulich

gemacht werden. Zu diesem Zweck werden auf der Basis einer Dunkelfeldbefragung von Schülern aus Niedersachsen aus dem Jahr 2013 Prädiktoren von Viktimisierungserfahrungen untersucht.

2 Einflussfaktoren von Viktimisierungserfahrungen Jugendlicher

Im Gegensatz zur Erklärung abweichenden Verhaltens existieren keine eigenständigen Theorien, die sich mit der Erklärung von Viktimisierung im Allgemeinen oder auch speziell von Jugendlichen beschäftigen. Jedoch lässt sich aus verschiedenen kriminologischen Theorien, wie etwa dem *Routine-Activity-Ansatz* (Cohen/Felson 1979), der *General Theory of Crime* (Gottfredson/Hirschi 1990) und dem Ansatz der sozialen Desorganisation (Shaw/McKay 1969; Sampson u. a. 1997) sowie aus der bisherigen Forschung eine Reihe von Prädiktoren von Viktimisierung ableiten.

So zeigen Studien (u. a. Baier/Prätor im Druck; Gruszczynska u. a. 2012), dass Jungen ein höheres Risiko für Viktimisierung durch Gewalt- und Eigentumsdelikte, Mädchen dagegen ein höheres Viktimisierungsrisiko für sexuelle Gewalt aufweisen. Der *Migrationshintergrund* beeinflusst ebenfalls das Risiko der Opfererfahrung. Neben rassistischen Delikten und Diskriminierung, die nur von Personen mit Migrationshintergrund erlebt werden können, gilt dies auch für andere Delikte. So finden etwa Baier und Prätor (im Druck) etwas höhere Prävalenzraten für Viktimisierung durch Gewaltdelikte von Jugendlichen mit Migrationshintergrund verglichen mit deutschen Jugendlichen. Allerdings variiert die Prävalenz stark über verschiedene Migrantengruppen hinweg.

In der empirischen Forschung hat sich zudem gezeigt, dass *intensiveres elterliches Monitoring* das Risiko für Diebstahl reduziert (Gruszczynska u. a. 2012). Als wichtiger Prädiktor hat sich auch *elterliche Gewalt* erwiesen: Jugendliche, die vor ihrem zwölften Lebensjahr schwere elterliche Gewalt erlebt haben, werden häufiger Opfer von Gewalt (Baier/Prätor im Druck).

Das *Freizeitverhalten* Jugendlicher spielt ebenfalls eine Rolle. Baier und Prätor (im Druck) berichten, dass Schülerinnen und Schüler, die Zeit in Bars, Diskotheken etc. verbringen, ein höheres Risiko aufweisen, Opfer von Gewaltdelikten zu werden. Auch der Kontakt mit *delinquenten Freundinnen und Freunden* erhöht das Viktimisierungsrisiko. Jugendliche mit delinquenten Freundinnen und Freunde werden durch diese in risikoreiche Aktivitäten verwickelt, die etwa mit gewalttätigen Auseinandersetzungen einhergehen und auch das Viktimisierungsrisiko für andere Delikte erhöhen können. Empirische Studien konnten signifikant positive Effekte sowohl für Diebstahl, Körperverletzung und Raub (Gruszczynska u. a. 2012) als auch für Gewaltdelikte (Baier/Prätor im Druck) insgesamt zeigen.

Jugendliche verbringen große Teile ihre Zeit in der *Schule*. Aus dem Desorganisationsansatz, der sich auch auf die Schule übertragen lässt (Hanslmaier 2014), kann abgeleitet werden, dass das Viktimisierungsrisiko von der Fähigkeit des Kontexts Schule abhängt, informelle soziale Kontrolle auszuüben. So zeigt etwa die Arbeit von Sapouna (2010), dass Viktimisierung durch Bullying in Klassen mit höherer kollektiver Wirksamkeit schwächer ist. Zudem berichten Gruszczynska u. a. (2012), dass Schülerinnen und Schüler in desorganisierten Schulen ein höheres Risiko aufweisen, Opfer von Diebstahl, Körperverletzung oder Raub zu werden.

3 Datensatz und Operationalisierung

Die empirischen Analysen im vorliegenden Beitrag basieren auf dem Niedersachsensurvey 2013. Dieser Datensatz erlaubt die Analyse einer großen Bandbreite verschiedener Prädiktoren der Inzidenz und Prävalenz von Viktimisierung und weist zudem eine hierarchische Datenstruktur auf. Der Niedersachsensurvey 2013¹ wurde im Jahr 2013 vom Kriminologischen Forschungsinstitut Niedersachsen durchgeführt und ist repräsentativ für die 9. Jahrgangsstufe in Niedersachsen. Für die Stichprobe wurden zunächst aus einer Liste aller Schulklassen in Niedersachsen (geschichtet nach sieben Schultypen) die zu befragenden Klassen gezogen. Alle am Befragungstag anwesenden Schülerinnen und Schüler der ausgewählten Klassen wurden schriftlich und anonym im Klassenverband im Rahmen des Schulunterrichts im Beisein einer/eines geschulten Testleiterin bzw. Testleiters und einer Lehrperson befragt. Dementsprechend stellt die Stichprobe eine Klumpenstichprobe dar und die Daten weisen eine hierarchische Struktur auf. Die Rücklaufquote betrug 64,4 %, sodass die Stichprobe 9.512 Schülerinnen und Schüler aus 485 Klassen umfasst. Die Auswertungen in diesem Beitrag beschränken sich auf die 8.411 Schüler, die für alle im Folgenden betrachteten Variablen gültige Werte aufweisen.²

Für die Auswertungen werden drei Indikatoren der Viktimisierung herangezogen: Die *Prävalenz von Gewaltdelikten* gibt an, ob die Jugendlichen in den letzten zwölf Monaten Opfer mindestens eines von sechs Delikten (Raub, Erpressung, sexuelle Gewalt, Körperverletzung mit Waffen, Körperverletzung durch einzelne Personen und Körperverletzung durch mehrere Personen) geworden sind (siehe *Tabelle 1* für deskriptive Statistiken aller Variablen). Die

¹ Für eine genauere Beschreibung der Studie siehe Baier 2015, 28 ff.

² Insbesondere Befragte aus Förderschulen fallen dadurch aus der Stichprobe, da diese eine kürzere Version des Fragebogens beantwortet haben, die einige der betrachteten Variablen nicht enthält.

Inzidenz für Eigentumsdelikte (zwölf Monate) ergab sich aus der Summe der Opfererfahrungen für vier Delikte (Fahrraddiebstahl, anderer Fahrzeugdiebstahl, Diebstahl und Sachbeschädigung). *Opfer von Schulgewalt* sind Befragte, die mindestens einmal im letzten Schulhalbjahr mindestens eine von zwei Formen von Gewalt erlebt haben („Ich wurde von anderen Schülern absichtlich geschlagen oder getreten“ bzw. „Andere Schüler haben mich erpresst und gezwungen, Geld oder Sachen herzugeben“).

Ein *Migrationshintergrund* lag vor, wenn die/der Befragte oder deren/dessen leibliche Mutter oder leiblicher Vater eine andere als die deutsche Staatsangehörigkeit besaß oder nicht in Deutschland geboren war. Die sieben Schultypen wurden zu drei *Schulformen* zusammengefasst: Hauptschule, Realschule (inkl. integrierte Haupt- und Realschule) und Gymnasium (inkl. Gesamtschule).

Selbstkontrolle wurde über die Subdimension *Risikosuche* mit vier Items (Beispielitem „Ich teste gerne meine Grenzen, indem ich etwas Gefährliches mache“) auf einer vierfach gestuften Skala erhoben (Cronbachs $\alpha = 0,859$).

Das *elterliche Monitoring* wurde über drei Items zu elterlichem Erziehungsverhalten vor dem zwölften Lebensjahr erfasst (Beispielitem Mutter/Vater hat „genau gewusst, wo ich in meiner Freizeit bin“). Die Befragten machten jeweils getrennte Angaben für Mutter und Vater, die zunächst gemittelt wurden bevor der Mittelwert aus den drei Items gebildet wurde. Hohe Werte geben ein intensives elterliches Monitoring an (Cronbachs $\alpha = 0,693$).

Überdies wurde die erlebte *elterliche Gewalt* vor dem zwölften Lebensjahr erfasst. Die Schülerinnen und Schüler beantworteten sechs Items zur Häufigkeit erlebter Gewalt getrennt für Mutter und Vater auf einer sechsstufigen Skala. Zunächst wurde für jedes Item der Maximalwert aus den Angaben für Mutter und Vater und daraus anschließend drei Kategorien gebildet. Jugendliche, die keine der sechs Gewaltformen erlebt hatten, wurden als „keine Gewalt“ klassifiziert, wer mindestens eine der leichteren Gewaltformen („mir eine runtergehauen“/„mich hart angepackt oder gestoßen“/„mit einem Gegenstand nach mir geworfen“), aber keine der drei schwereren Gewaltformen („mich mit einem Gegenstand geschlagen“/„mich mit der Faust geschlagen oder mich getreten“/„mich geprügelt, zusammengeschlagen“) erlebt hatte, wurde mit „leichte Gewalt“ bezeichnet. In der Gruppe „schwere Gewalt“ sind Jugendliche, die mindestens eine der drei schwereren Gewaltformen erlebt haben.

Für sechs abweichende Verhaltensweisen (Ladendiebstahl, Schwänzen, Raub, Körperverletzung, Sachbeschädigung und Verkauf von Drogen) wurde getrennt erhoben, wie viele Freundinnen bzw. Freunde die Jugendlichen haben, die diese Verhaltensweisen in den letzten zwölf Monaten ausgeführt haben.

Die *Anzahl der delinquenten* Freundinnen und Freunde wurde aus diesen Angaben mittels Maximalwertbefehl (höchster Wert aus allen sechs Angaben) gebildet und in drei Kategorien eingeteilt.

Die *Routineaktivitäten der Jugendlichen* wurden über die Zeit erfasst, die diese pro Woche damit verbringen, „in Kneipe, Disco, Kino, zu Veranstaltungen“ zu gehen. Die Jugendlichen wurden dann in vier Gruppen anhand der empirischen Verteilung der Variable eingeteilt. Die erste Gruppe umfasst die Jugendlichen, die keine Zeit in Kneipen etc. verbringen, die weiteren Grenzen waren der Median und das dritte Quartil.

Darüber hinaus wurden Eigenschaften der Schule erhoben. Diese stellen Kollektivmerkmale dar, da es sich um Eigenschaften des Kontexts Schule handelt. Für diese Merkmale werden die Angaben der einzelnen Schülerinnen und Schüler aus derselben Klasse aggregiert, d. h. der Mittelwert über alle Schüler einer Klasse gebildet.³ Einschränkend ist an dieser Stelle anzumerken, dass es sich bei den Eigenschaften um Eigenschaften der Klasse und nicht der Schule handelt. Dies ist auch der Datenstruktur geschuldet, da 485 Klassen aus 379 Schulen befragt wurden, d. h. zumeist nur eine Klasse pro Schule. *Schulische Desorganisation* wurde mit zwei Items erhoben (Beispielitem „An meiner Schule gibt es viel Gewalt“; $r=0,551$), die die Schüler auf einer vierfach gestuften Skala von „stimmt nicht“ bis „stimmt genau“ bewerten sollten. Auf der gleichen Skala wurden acht Items verwendet, um *schulische Kohäsion* zu erfassen (Beispielitem „Wir halten in meiner Klasse fest zusammen“; Cronbachs $\alpha=0,790$). Für beide Variablen wurde zunächst der Mittelwert über die jeweiligen Items für jeden Befragten gebildet und dann auf der Ebene der Klasse aggregiert, wobei mindestens vier gültige Individualangaben pro Klasse zur Verfügung standen.

³ Für eine weiterführende Diskussion der Bildung von Kontexteigenschaften aus aggregierten Individualangaben v. a. im Hinblick auf Validität und Reliabilität siehe u. a. Oberwittler 2003.

Tabelle 1:

Deskriptive Statistiken der Variablen

		MW	SD	Min	Max
Individualvariablen (Level-1) n = 8.411 Schüler					
Prävalenz Gewaltdelikte	0 „Nichtopfer“ 1 „Opfer“	0,13	0,33	0	1
Inzidenz Eigentumsdelikte	Anzahl der Opfererfahrungen	0,59	2,78	0	120
Prävalenz für Schulgewalt	0 „Nichtopfer“ 1 „Opfer“	0,18	0,38	0	1
Geschlecht	0 „weiblich“ 1 „männlich“	0,50	0,50	0	1
Migrationshintergrund	0 „kein Migrationshintergrund“ 1 „Migrationshintergrund“	0,24	0,42	0	1
Schulform	„Hauptschule“	0,07	0,25	0	1
<i>Dummyvariablen</i>	„Realschule“	0,45	0,50	0	1
	„Gymnasium/Gesamtschule“	0,48	0,50	0	1
Risikosuche	Mittelwertskala aus vier Items	2,10	0,76	1	4
elterliches Monitoring	Mittelwertskala aus drei Items	4,02	0,72	1	5
elterliche Gewalt	„keine Gewalt“	0,57	0,50	0	1
<i>Dummyvariablen</i>	„leichte Gewalt“	0,31	0,46	0	1
	„schwere Gewalt“	0,12	0,33	0	1
delinquente Freunde	„keine delinquenten Freunde“	0,36	0,48	0	1
<i>Dummyvariablen</i>	„1 bis 5 delinquente Freunde“	0,55	0,50	0	1
	„6 und mehr delinquente Freunde“	0,09	0,29	0	1
Zeit pro Woche in Kneipe, Disco, Kino etc.	„0 Minuten pro Woche“	0,39	0,49	0	1
<i>Dummyvariablen</i>	„1 bis unter 240 Minuten pro Woche“	0,10	0,29	0	1
	„240 bis unter 480 Minuten“	0,24	0,43	0	1
	„480 und mehr Minuten“	0,27	0,45	0	1
Eigenschaften auf Klassenebene (Level-2) n = 454 Klassen					
Desorganisation	Mittelwertskala aus zwei Items	2,09	0,34	1,32	3,22
Kohäsion	Mittelwertskala aus acht Items	2,67	0,24	1,90	3,27

4 Empirische Analysen und Ergebnisse

Bevor die logistische Regression sowie Modelle für Zähl- und Mehrebenen-daten vorgestellt werden, soll das einfache lineare Regressionsmodell (OLS) rekapituliert werden, da dies als Ausgangspunkt für die Erklärung der weiteren Verfahren dient.⁴ Ziel des linearen Regressionsmodells ist es, eine abhängige Variable y durch eine oder – im multivariaten Fall – k unabhängige Variablen x_j (mit $j = 1, \dots, k$) zu erklären. Als Gleichung lässt sich der Wert der abhängigen Variable y_i für die Beobachtung i darstellen als:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (1)$$

Mit x_{1i} und x_{2i} sind die Werte der (in diesem Fall zwei) unabhängigen Variablen für die Beobachtung i bezeichnet. Der Koeffizient β_0 bezeichnet den y -Achsenabschnitt (Intercept), die Koeffizienten (Slopes) β_1 und β_2 geben jeweils den Zusammenhang zwischen den Ausprägungen (genauer: den linearen Einfluss) der unabhängigen Variable x_1 bzw. x_2 und der abhängigen Variable an. Der Fehlerterm ε_i beinhaltet alle nicht im Modell enthaltenen Einflüsse und wird als zufällig und normalverteilt mit einem Mittelwert von null angenommen. Wenn dies zutrifft, kann dieser zufällige Faktor vernachlässigt werden und der mit \hat{y}_i bezeichnete geschätzte Wert der abhängigen Variable der Beobachtung i (z. B. Einkommen) ergibt sich aus den Werten des Befragten für x_1 (z. B. Bildung) und x_2 (z. B. berufliche Stellung) und den entsprechenden Werten der drei Koeffizienten. Die Abweichung zwischen dem tatsächlichen Wert y_i und dem geschätzten Wert \hat{y}_i wird als Residuum $\varepsilon_i = y_i - \hat{y}_i$ ausgedrückt. Für die Schätzung der Koeffizienten aus Gleichung (1) werden diese so gewählt, dass die quadrierten Residuen minimal sind (Methode der kleinsten Quadrate = *Ordinary Least Squares* = OLS).

4.1 Logistische Regression: Analyse von Prävalenzdaten

4.1.1 Lineares Wahrscheinlichkeitsmodell und logistische Regression

Daten zur Prävalenz von Opferschaft nehmen *qua definitione* nur zwei Zustände an: Opfer und Nichtopfer. Derartige Merkmale werden in den Sozialwissenschaften als binäre oder dichotome Merkmale bezeichnet und für die statistischen Analysen mit 0 und 1 codiert. Im vorliegenden Fall würde man Nichtopfer mit 0 und Opfer mit 1 codieren und versuchen, die Varianz dieser

⁴ Die Darstellung lehnt sich an Windzio 2013, 17–21 an.

Variable mithilfe der unabhängigen Variablen in einem multivariaten Modell zu erklären.

Generell besteht die Möglichkeit, ein Modell zur Erklärung von Prävalenzen in Abhängigkeit diverser unabhängiger Variablen mittels OLS-Regression zu schätzen. Dieses Verfahren wird als lineares Wahrscheinlichkeitsmodell (*Linear Probability Model* = LPM) bezeichnet. Die tatsächliche abhängige Variable ist in diesem Fall allerdings nicht binär, sondern metrisch und wird als Wahrscheinlichkeit $P(Y = 1)$ dessen interpretiert, dass die abhängige Variable den Wert 1 annimmt (Best/Wolf 2010, 828; Windzio 2013, 39–40).

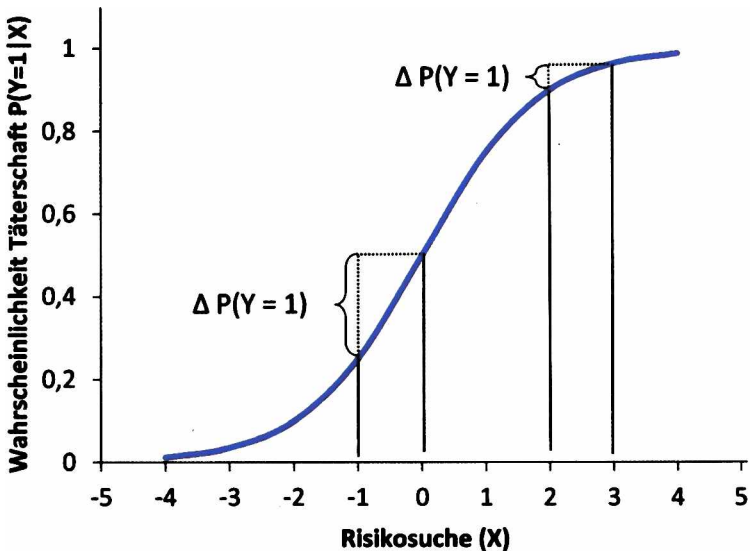
Allerdings führt das LPM zu einer Reihe von Problemen (Best/Wolf 2010, 830; Windzio 2013, 40–42; u. a. Long 1997, 38–40):

- Die durch das Regressionsmodell vorhergesagten Werte sind möglicherweise kleiner als 0 oder größer als 1 und liegen somit außerhalb des für Wahrscheinlichkeiten definierten Bereichs.
- Da die abhängige Variable nur die Werte 0 und 1 annehmen kann, sind die Residuen heteroskedastisch. Heteroskedastizität bedeutet, dass die Varianz der Residuen systematisch von den Werten der unabhängigen Variable abhängt. Dies führt zu verzerrten Standardfehlern und beeinflusst somit die Inferenzstatistiken der Koeffizienten, die Rückschlüsse über die Existenz von Effekten in der Grundgesamtheit erlauben.
- Die Normalverteilungsannahme der Residuen ist verletzt. Diese können für jede Konstellation der unabhängigen Variablen nur zwei Werte annehmen.
- Darüber hinaus stellt sich die Frage, ob der lineare Zusammenhang zwischen den unabhängigen Variablen und der abhängigen Variable, von dem das LPM ausgeht, funktional angemessen ist. So ist davon auszugehen, dass sich die Wahrscheinlichkeiten den Werten 0 und 1 in Abhängigkeit des Prädiktors nicht linear, sondern asymptotisch annähern, d. h., die Steigung der Funktion, die die Wahrscheinlichkeit von $Y = 1$ in Abhängigkeit der Ausprägungen der Prädiktoren angibt, sollte am unteren und oberen Bereich geringer als in der Mitte ausfallen. Das bedeutet, dass sich der Effekt eines zusätzlichen Anstiegs der unabhängigen Variable reduziert, wenn sich die Wahrscheinlichkeit den Randbereichen 0 und 1 annähert. Geht man beispielsweise davon aus, dass die Wahrscheinlichkeit eines jugendlichen, Gewalttäter zu werden, vom Ausmaß seiner Risikobereitschaft abhängt, dann wird der Effekt einer Veränderung der Risikobereitschaft um einen bestimmten Betrag für Jugendliche mit einer sehr geringen oder sehr hohen Risikobereitschaft nur geringen Einfluss auf de-

ren Täterwahrscheinlichkeit haben, da diese ohnehin bereits sehr hoch (hohe Risikobereitschaft) oder sehr niedrig ist (niedrige Risikobereitschaft). Bei Jugendlichen mit mittlerem Niveau an Risikobereitschaft kann deren Veränderung um den gleichen Betrag einen größeren Einfluss auf die Täterwahrscheinlichkeit haben. *Abbildung 1* zeigt dies exemplarisch. So ist der Effekt des Anstiegs der Risikosuche um eine Einheit auf die Veränderung der Täterwahrscheinlichkeit $\Delta P(Y=1)$ davon abhängig, wo man sich auf der x-Achse befindet. Steigt die Risikosuche x von -1 auf 0 , wächst die Wahrscheinlichkeit stärker, als wenn x von 2 auf 3 steigt. In der Abbildung wurde die Skalierung so gewählt, dass die Risikosuche einen Wertebereich von -4 bis 4 hat. Somit entspricht der Anstieg von -1 auf 0 einem Anstieg im mittleren Wertebereich, während der Anstieg von 2 auf 3 sich im höheren Wertebereich abspielt.

Abbildung 1:

Effekt des Anstiegs von X um eine Einheit auf die Veränderung der Wahrscheinlichkeit von Y = 1 in Abhängigkeit von X



Die logistische Regression kann diese Probleme des LPM lösen. Zu diesem Zweck wird die Schätzgleichung transformiert. Die erste Umformung wandelt die Wahrscheinlichkeit, dass Y den Wert 1 annimmt, also $P(Y=1)$, in Odds (= Chancen) um. Odds stellen Verhältnisse von Wahrscheinlichkeiten

dar – genauer die Beziehung von Eintrittswahrscheinlichkeit zu Gegenwahrscheinlichkeit:

$$\text{Odds} = O = \frac{P(Y=1)}{1-P(Y=1)} \quad (2)$$

Liegt etwa die Wahrscheinlichkeit, Opfer einer Straftat zu werden, bei 20 %, dann entspricht dies Odds von $^{20}/_{80} = 0,25$. Für eine Wahrscheinlichkeit von 50 % liegen die Odds bei $^{50}/_{50} = 1$; für eine Wahrscheinlichkeit von 80 % liegen die Odds bei $^{80}/_{20} = 4$. Odds haben einen Wertebereich von 0 bis $+\infty$; allerdings ist zu beachten, dass die Transformation von Wahrscheinlichkeiten in Odds nicht linear verläuft. Im Weiteren wird von Odds auch als Chance oder Risiko gesprochen; wenn es um Wahrscheinlichkeiten geht, wird der Begriff Wahrscheinlichkeit verwendet. Diese begriffliche Festlegung ist notwendig, da die Begriffe Risiko bzw. Chance im Deutschen nicht ganz eindeutig sind.

Um die Begrenzung der Odds nach unten ebenfalls zu eliminieren, werden sie logarithmiert. Diese sogenannten Logits haben einen Wertebereich von $-\infty$ bis $+\infty$, da Odds zwischen 0 und 1 durch das Logarithmieren negativ werden (Best/Wolf 2010, 829; Kohler/Kreuter 2008, 262–267; Long 1997, 50–54).

Die durch diese zweifache Transformation aus den Wahrscheinlichkeiten berechneten Logits dienen dann als abhängige Variable einer Linearkombination der k unabhängigen Variablen x_i . Diese Linearkombination ist dem linearen Wahrscheinlichkeitsmodell (LPM) zwar ähnlich, wenngleich nicht die Wahrscheinlichkeit selbst, sondern deren Logit linear von den unabhängigen Variablen abhängt. Für die Interpretation der Ergebnisse bedeutet dies, dass eine Erhöhung einer unabhängigen Variable x_i um eine Einheit nicht die Wahrscheinlichkeit, sondern die logarithmierten Odds der Wahrscheinlichkeit $P(Y=1)$ um β_i Einheiten ändert. Aufgrund dieser Transformationen ist der Zusammenhang zwischen den unabhängigen Variablen und der Wahrscheinlichkeit nicht linear. Dies bringt Herausforderungen für die Interpretation mit sich, auf die im nächsten Abschnitt eingegangen wird. Durch Transformationen der Gleichung wird jedoch erreicht, dass die abhängige Variable einen nicht begrenzten Wertebereich hat und die funktionale Form des Zusammenhangs zwischen den Prädiktoren und der Auftrittswahrscheinlichkeit eine sinnvolle Form⁵ annimmt (Best/Wolf 2010, 829), wie im vorangegangenen Beispiel für den Zusammenhang zwischen Risikosuche und Täterschaft ausgeführt. Die Schätzung dieser Gleichung erfolgt im Gegensatz zum linearen Wahrscheinlichkeitsmodell (LPM) nicht mit der Methode der kleinsten Qua-

⁵ Für eine ausführlichere Diskussion hierzu siehe Windzio 2013, 39–47.

drate (OLS) sondern mittels Maximum-Likelihood⁶ (Best/Wolf 2010, 830; Kohler/Kreuter 2008, 267).⁷

4.1.2 Durchführung und Interpretation der logistischen Regression

Durchführung und Interpretation der logistischen Regression sollen am Beispiel der Prävalenz von Gewaltkriminalität dargestellt werden. Für diese und alle weiteren Analysen in diesem Beitrag wurde auf die Statistiksoftware Stata zurückgegriffen. Im Weiteren wird in Fußnoten jeweils auf die konkreten Stata-Befehle hingewiesen.

In *Tabelle 2* ist das Ergebnis einer logistischen Regression⁸ für die Prävalenz der Opferschaft von Gewaltdelikten dargestellt. In der ersten Spalte sind die Koeffizienten der logistischen Regression als Logits dargestellt. Diese können analog zur linearen Regression interpretiert werden: Ein Anstieg einer unabhängigen Variablen (z. B. der Risikosuche) um eine Einheit erhöht die Logits (also die logarithmierten Odds), Opfer einer Gewalttat zu werden, um 0,271. Analog lassen sich Dummyvariablen interpretieren: Hier gibt der jeweilige Koeffizient den Unterschied der Logits zur Referenzgruppe an. Positive Koeffizienten stehen somit für einen positiven Effekt, negative Koeffizienten für einen negativen Effekt des jeweiligen Prädiktors auf die abhängige Variable. Allerdings lassen sich die Logits aufgrund der nicht linearen Transformation der Wahrscheinlichkeiten nicht inhaltlich, sondern nur hinsichtlich Vorzeichen und Signifikanz interpretieren (Best/Wolf 2010, 831).

Die abgebildeten Logits aus *Tabelle 2* (Spalte 1) zeigen ein bei Hauptschülerinnen und Hauptschülern höheres Risiko für Gewaltviktimsierung als bei Realschülerinnen und Realschülern und Gymnasiastinnen und Gymnasiasten, wobei nur der Unterschied zu Letzteren signifikant ist. Ein wichtiger Einflussfaktor ist daneben das Elternhaus: Elterliches Monitoring reduziert das Viktimisierungsrisiko signifikant. Demgegenüber erhöht in der Kindheitspha-

⁶ Bei der OLS-Regression werden die Parameter geschätzt, indem die quadrierten Abweichungen zwischen den vorhergesagten und den beobachteten Werten minimiert werden. Beim Maximum-Likelihood-Verfahren hingegen wird eine Annahme über die durch die unabhängigen Variablen bedingte Verteilung der abhängigen Variable getroffen. Dann werden die unbekannt Parameter β , so geschätzt, dass die Wahrscheinlichkeit, die beobachteten Werte zu erhalten, maximal ist (Kohler/Kreuter 2008, 267–271; ausführlich Gautschi 2010).

⁷ Das logistische Regressionsmodell lässt sich auch als latentes Variablenmodell herleiten. Da die Herleitung über die Transformation von Wahrscheinlichkeiten aber besser verständlich ist, wurde darauf verzichtet. Für eine ausführlichere Darstellung siehe Long 1997, 40–50. Kürzere Darstellungen finden sich in Long und Freese 2003, 110–113 sowie Best und Wolf 2010, 834–836.

⁸ Zu diesem Zweck wurde der Stata-Befehl *logit* verwendet. Die Ergebnisse lassen sich über die entsprechende Option des Befehls entweder als Logits oder als Odds Ratios ausgeben.

se bis zum Alter von 12 Jahren erlebte elterliche Gewalt das Risiko der Jugendlichen, Opfer einer Gewalttat zu werden. Ein weiterer wichtiger Prädiktor ist das Freizeitverhalten: Jugendliche, die delinquente Freunde haben und viel Zeit in Kneipen, Discotheken etc. verbringen, haben ein erhöhtes Risiko, Opfer von Gewalt zu werden.

Tabelle 2:

Logistische Regression der Prävalenz für Gewaltdelikte

	Logit	OR	AME
männlich (1 = ja)	0,042	1,043	0,004
Migrationshintergrund (1 = ja)	0,003	1,003	0,000
Schulform			
Hauptschule	Referenz	Referenz	Referenz
Realschule	-0,182	0,833	-0,021
Gymnasium	-0,470***	0,625***	-0,050***
Risikosuche	0,271***	1,311***	0,027***
elterliches Monitoring	-0,232***	0,793***	-0,023***
elterliche Gewalt			
keine Gewalt	Referenz	Referenz	Referenz
leichte Gewalt	0,725***	2,065***	0,070***
schwere Gewalt	1,275***	3,580***	0,150***
delinquente Freunde			
keine delinquenten Freunde	Referenz	Referenz	Referenz
1 bis 5 delinquente Freunde	0,537***	1,711***	0,049***
6 und mehr delinquente Freunde	1,153***	3,166***	0,130***
Freizeitverhalten			
keine Zeit in Kneipe etc.	Referenz	Referenz	Referenz
wenig Zeit in Kneipe etc.	-0,032	0,969	-0,003
moderat Zeit in Kneipe etc.	0,114	1,121	0,011
viel Zeit in Kneipe etc.	0,269**	1,309**	0,028**
Konstante	-2,403***	0,090***	
Pseudo R ²	0,102		

n = 8.411; *** p < 0,001, ** p < 0,01, * p < 0,05

Eine größere inhaltliche Aussagekraft der Koeffizienten ergibt sich durch eine Umwandlung der Logits durch Entlogarithmierung in *Odds Ratios* (OR). OR sind Verhältnisse von Odds; für die Interpretation von Dummyvariablen bedeutet dies im vorliegenden Beispiel etwa, dass die Chance bzw. das Risiko (Odds), Opfer einer Gewalttat zu werden, für Jungen 1,043-mal so groß ist wie für Mädchen (nicht signifikant). Dabei gilt es zu beachten, dass sich die OR auf die Verhältnisse von Odds und nicht von Wahrscheinlichkeiten beziehen.

Bei der Interpretation der Odds Ratios kontinuierlicher Variablen ist die multiplikative Verknüpfung zwischen der unabhängigen Variablen und den Odds zu beachten. So erhöht der Anstieg der Risikosuche um eine Einheit das Risiko, Opfer einer Straftat zu werden auf das 1,311-Fache; ein Anstieg der Risikosuche um zwei Einheiten hingegen steigert das Risiko auf das $1,311 \times 1,311 = 1,719$ -Fache. OR kleiner als 1 geben einen negativen Effekt an, OR größer als 1 einen positiven Effekt (Kohler/Kreuter 2008, 274–275; Best/Wolf 2010, 831–832).

Allerdings ist auch die Interpretation der OR nur bedingt aussagekräftig, wie Best und Wolf (2010, 832–833) ausführen. So handelt es sich um Verhältnisse von Odds (= Wahrscheinlichkeitsverhältnisse). Ohne die Basiswahrscheinlichkeit (bzw. die Odds), als Mädchen Opfer von Gewalt zu werden, zu kennen, kann man keine Aussage über das absolute Risiko etwa der Jungen, Opfer von Gewalt zu werden, treffen. Auch ist die Frage nach der Bedeutung eines Effekts nicht auf Basis der OR zu klären, da z. B. eine Verdoppelung des Risikos je nach Höhe des Ausgangsrisikos einen unterschiedlich großen Effekt auf die Wahrscheinlichkeit hat (auch Windzio 2013, 53–54).

Anschaulicher als OR sind Interpretationen auf der Basis prognostizierter Wahrscheinlichkeiten. Wahrscheinlichkeiten sind darüber hinaus das eingangs dargestellte eigentliche Ziel der logistischen Regression. Möchte man den Effekt einer oder mehrerer unabhängiger Variablen auf die vorhergesagte Wahrscheinlichkeit darstellen, ergibt sich die Schwierigkeit, dass das Verhältnis von Logits und Wahrscheinlichkeiten nicht linear ist. Dies bewirkt, dass der Effekt einer Zunahme der unabhängigen Variable um eine Einheit auf die vorhergesagte Wahrscheinlichkeit über den Wertebereich der unabhängigen Variable variiert. Darüber hinaus hängt der Effekt des Anstiegs einer unabhängigen Variable um eine Einheit im multivariaten Modell auch von den Ausprägungen der übrigen unabhängigen Variablen ab (Windzio 2013, 62–64; auch Best/Wolf 2010; Kohler/Kreuter 2008).

Folgendes Beispiel soll den Effekt des Anstiegs einer unabhängigen Variablen auf die Wahrscheinlichkeit, z. B. Opfer einer Straftat zu werden, in Abhängigkeit des Basisrisikos bei gegebenem OR verdeutlichen. Hat beispielsweise eine Variable einen OR von zwei, verdoppelt sich das Risiko,

wenn die Variable um eine Einheit ansteigt. Liegt das Basisrisiko bei ${}^5/_{100}$, führt eine Verdoppelung des Risikos zu Odds von ${}^{10}/_{100}$. Dementsprechend steigt die Wahrscheinlichkeit von ${}^5/_{105}=0,048$ auf ${}^{10}/_{110}=0,091$. Würde das Basisrisiko bei ${}^{50}/_{50}=1$ liegen, würde die Wahrscheinlichkeit von ${}^{50}/_{100}=0,500$ auf ${}^{100}/_{150}=0,667$ steigen. Folglich erhöht der Anstieg der unabhängigen Variablen die Wahrscheinlichkeit je nach Basisrisiko um 0,043 bzw. 0,167. Das Basisrisiko wiederum hängt von den Ausprägungen der anderen Variablen ab.

Dementsprechend kann der Effekt einer unabhängigen Variablen auf die Wahrscheinlichkeit nicht ohne Weiteres in einer Zahl ausgedrückt werden. Gleichwohl bestehen diverse Möglichkeiten, um den Einfluss der unabhängigen Variablen auf die Wahrscheinlichkeit darzulegen.

Eine Möglichkeit ist die *grafische Darstellung*. Hierbei wird die vorhergesagte Wahrscheinlichkeit für verschiedene Ausprägungen einer unabhängigen Variablen berechnet (ggf. getrennt für verschiedene Subgruppen). Da die vorhergesagte Wahrscheinlichkeit auch von den Ausprägungen der anderen unabhängigen Variablen abhängt, muss für alle anderen Prädiktoren, also unabhängigen Variablen, ein Wert festgelegt werden.

Abbildung 2:

Vorhergesagte Wahrscheinlichkeiten

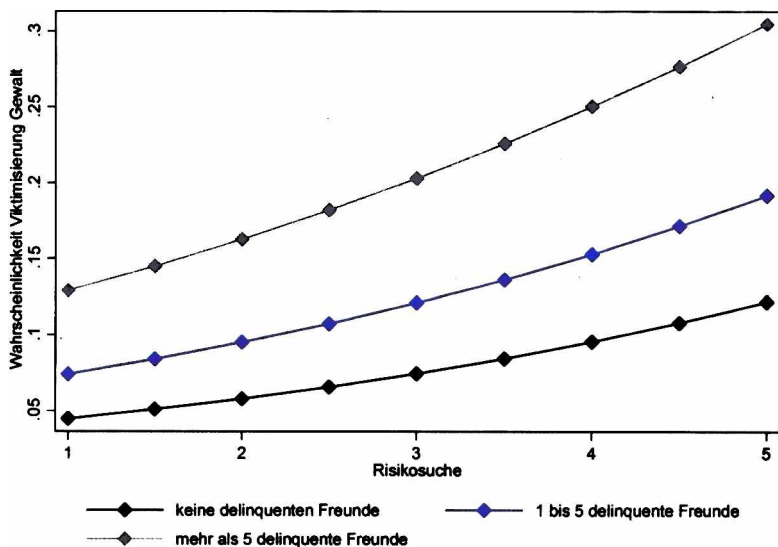


Abbildung 2⁹ stellt die vorhergesagte Wahrscheinlichkeit für Gewaltviktimsierung in Abhängigkeit von der Risikosuche und der Zahl der delinquenten Freunde dar. Alle anderen Prädiktoren aus *Tabelle 2* sind am Mittelwert (kontinuierliche Variablen) bzw. an der Referenzkategorie (Dummyvariablen) konstant gehalten. Es wird deutlich, dass die Wahrscheinlichkeit, Opfer von Gewalt zu werden, mit zunehmender Risikosuche steigt. Jedoch hängt der Betrag des Anstiegs davon ab, wie viele delinquente Freunde man hat und welchen Ausgangswert man zugrunde legt, d. h. ob die Risikosuche von 1 auf 2 oder von 4 auf 5 steigt. Dies verdeutlicht noch einmal grafisch die nicht linearen Zusammenhänge zwischen den Variablen und den vorhergesagten Wahrscheinlichkeiten.

Eine zweite Möglichkeit der Darstellung sind sogenannte *marginale Effekte*. Diese geben die Veränderung der Wahrscheinlichkeit an, wenn sich die betreffende Variable um einen infinitesimal kleinen (d. h. gegen null strebenden) Betrag ändert, d. h., es geht um die Steigung der logistischen Funktion (Windzio 2013, 65–66). Auch der marginale Effekt hängt von Ausprägungen aller Variablen im Modell ab (Long/Freese 2003, 139). Demensprechend lassen sich verschiedene marginale Effekte unterscheiden (Windzio 2013, 66–67).¹⁰ Der *Marginal Effect at the Mean* (MEM) berechnet den marginalen Effekt, wenn alle unabhängigen Variablen den Mittelwert aufweisen. Im Gegensatz dazu berechnet der *Average Marginal Effect* (AME)¹¹ den Mittelwert der marginalen Effekte aus allen im Modell enthaltenen Beobachtungen (Best/Wolf 2010; Long 1997, 74; Windzio 2013, 66). Die AME haben, wie Best und Wolf (2010, 840) ausführen, gegenüber den MEM den Vorteil, dass sie Koeffizienten von Modellen vergleichen können, in die schrittweise mehr Variablen aufgenommen werden. Die letzte Spalte in *Tabelle 2* zeigt die AME für die Prävalenz von Gewaltkriminalität.¹² Dabei tritt zutage, dass die Wahrscheinlichkeit, Opfer einer Gewalttat zu werden, durchschnittlich um 2,7 Prozentpunkte steigt, wenn die Risikosuche um eine Einheit zunimmt; betrachtet

⁹ Die vorhergesagten Wahrscheinlichkeiten, die die Basis der Grafik bilden, lassen sich mit dem Befehl *prgen* berechnen, der Teil des *spost9*-Pakets ist. Das Paket *spost9* wurde von J. Scott Long und Jeremy Freese programmiert und bietet eine Reihe hilfreicher Befehle für die Analyse kategorialer Daten. Dieses Zusatzpaket kann über den Befehl *findit spost9* installiert werden. Eine gute Darstellung des Paktes findet sich in Long und Freese 2003.

¹⁰ Für eine mathematische Herleitung der marginalen Effekte siehe Windzio 2013, 65–66 und ausführlicher Long 1997, 72–75.

¹¹ Die AME können mit dem Befehl *margins* ausgegeben werden. Dieser Befehl ist ab Version 12 implementiert. Alternativ kann der Befehl *margeff* (Bartus 2005) genutzt werden.

¹² Möchte man Aussagen über die Grundgesamt treffen, sollte dies bei der Berechnung der Inferenzstatistiken (Standardfehler etc.) der AME berücksichtigt werden. Dies ist in Stata als Option des *margins*-Befehls verfügbar (StataCorp 2013, 1172–1173). Für die vorliegende Darstellung wurde darauf verzichtet.

man den Einfluss der Schulform, so zeigt sich, dass Realschülerinnen und Realschüler eine um 2,1 Prozentpunkte (Gymnasiastinnen und Gymnasiasten 5,0 Prozentpunkte) geringere Wahrscheinlichkeit aufweisen, Opfer zu werden, als Hauptschülerinnen und Hauptschüler (= Referenzkategorie). Den größten Effekt auf die Wahrscheinlichkeit, Opfer einer Gewalttat zu werden, haben elterliche Gewalt und der Kontakt zu delinquenten Freunden.

4.1.3 Modellfit, Modellvergleich und Mediationsanalysen

In der linearen Regression steht mit R^2 , also dem Anteil der durch das Modell erklärten Varianz an der Gesamtvarianz der abhängigen Variable, ein Maß für die Anpassungsgüte des Modells zur Verfügung. Für die logistische Regression existieren verschiedene Kennzahlen, die die Anpassungsgüte eines Modells angeben und zum Vergleich der Erklärungskraft von Modellen herangezogen werden können. Diese Maße sollten einen Wertebereich zwischen 0 und 1¹³ aufweisen, wobei ein Modell mit einem Wert von 0 keinerlei Erklärungskraft hat und Modelle, die höhere Werte erzielen eine größere Erklärungskraft besitzen.

Ein Maß, das auf dem Vergleich der Likelihood des geschätzten Modells mit der Likelihood eines Modells ohne erklärende Variablen (Nullmodell) basiert, ist *McFaddens Pseudo- R^2* . Andere Maße wie *Cox&Snell R^2* , *Cragg&Uhlen R^2* oder *Nagelkerke- R^2* basieren auf demselben Prinzip, enthalten aber zusätzlich eine Normierungs- oder Korrekturkomponente.¹⁴ Bei der Verwendung der Maße muss beachtet werden, dass es 1.) keinen einheitlichen Standard gibt, 2.) mehr unabhängige Variablen prinzipiell zu höheren Werten führen und 3.) verschiedene Maße nicht vergleichbar sind. So liefert Nagelkerke- R^2 immer höhere Werte als andere Maße (Best/Wolf 2010, 843–844).¹⁵

In multivariaten Modellen ist man oft daran interessiert, Mediationseffekte zu untersuchen, d. h. man beschäftigt sich mit der Frage, wie sich Koeffizienten verändern, wenn man weitere Variablen in das Modell aufnimmt. Das Problem in nicht linearen Modellen ist, dass die Regressionskoeffizienten nicht

¹³ Allerdings können nicht alle Maße (z. B. *McFaddens Pseudo R^2* , *Cox&Snell R^2*) auch tatsächlich den Wert 1 erreichen (Best/Wolf 2010, 843).

¹⁴ Stata gibt standardmäßig *McFaddens Pseudo- R^2* aus; der Befehl *fitstat* aus dem *spost9*-Paket berechnet zahlreiche weitere Fitmaße.

¹⁵ Neben diesen Maßen der Anpassungsgüte können auch Informationsmaße wie *Akaike's Information Criterion (AIC)* und das *Bayesian Information Criterion (BIC)* zur Modellselektion herangezogen werden (Best/Wolf 2010, 844; Long/Freese 2003, 94–95).

miteinander über Modelle hinweg verglichen werden können (Best/Wolf 2010, 838; Long 1997, 70; Mood 2009; Windzio 2013, 69–71 dort auch genauere Diskussion). Um dennoch Koeffizienten zwischen Modellen vergleichen zu können, bestehen zwei Möglichkeiten: Zum einen können wie bereits erwähnt die AME über Modelle hinweg verglichen, zum anderen voll- oder teilstandardisierte Koeffizienten berechnet werden, die Verzerrungen reduzieren, aber nicht komplett eliminieren (Best/Wolf 2010, 838–839).¹⁶

In *Tabelle 3* sind zwei Modelle für die Prävalenz von Gewaltdelikten dargestellt. Der Unterschied zwischen beiden Modellen besteht darin, dass in Modell 2 der Einfluss des elterlichen Erziehungsverhaltens hinzugefügt wurde (Modell 2 ist somit identisch mit dem Modell aus *Tabelle 2*). Betrachtet man zunächst den Modellfit, zeigt sich ein Anstieg der Erklärungskraft des Modells. Ein Likelihood-Ratio-Test¹⁷ ($LR \chi^2_{df=3} = 250,56; p < 0,001$) verdeutlicht zudem, dass die neu hinzugenommenen Parameter zu einer signifikanten Modellverbesserung führen.

¹⁶ Eine andere Möglichkeit ist die Reskalierung von Koeffizienten (Karlson u. a. 2012).

¹⁷ Ein Likelihood-Ratio-Test kann in Stata mit dem Befehl *lrtest* durchgeführt werden.

Tabelle 3:

Logistische Regression der Prävalenz für Gewaltdelikte: unstandardisierte und standardisierte Koeffizienten

Modell	(1)		(2)	
	Logit	stdXY	Logit	stdXY
männlich (1 = ja)	-0,008	-0,002	0,042	0,010
Migrationshintergrund (1 = ja)	0,174*	0,038	0,003	0,001
Schulform				
Hauptschule	Referenz		Referenz	
Realschule	-0,211	-0,055	-0,182	-0,045
Gymnasium	-0,548***	-0,142	-0,470***	-0,118
Risikosuche	0,345***	0,135	0,271***	0,103
delinquente Freunde				
keine delinquenten Freunde	Referenz			
1 bis 5 delinquente Freunde	0,673***	0,174	0,537***	0,134
6 und mehr delinquente Freunde	1,365***	0,203	1,153***	0,165
Freizeitverhalten				
keine Zeit in Kneipe etc.	Referenz		Referenz	
wenig Zeit in Kneipe etc.	-0,050	-0,008	-0,032	-0,005
moderat Zeit in Kneipe etc.	0,086	0,019	0,114	0,024
viel Zeit in Kneipe etc.	0,237**	0,055	0,269**	0,060
elterliches Monitoring			-0,232***	-0,083
elterliche Gewalt				
keine Gewalt			Referenz	
leichte Gewalt			0,725***	0,168
schwere Gewalt			1,275***	0,211
McFaddens R ²	0,063		0,102	

n = 8.411; *** p < 0,001, ** p < 0,01, * p < 0,05

Zusätzlich sind in *Tabelle 3* die standardisierten Koeffizienten aufgeführt.¹⁸ Standardisierte Koeffizienten β^s geben die Veränderung der Logits um β^s Standardabweichungen an, wenn die entsprechende unabhängige Variable x um eine Standardabweichung verändert wird, und erlauben einen modellübergreifenden Vergleich der Koeffizienten. Im vorliegenden Beispiel haben die

¹⁸ Diese lassen sich mit dem Befehl *listcoef* im Anschluss an die Modellschätzung ausgeben, der ebenfalls zum *post9*-Paket gehört.

meisten Variablen durch die Hinzunahme der weiteren Prädiktoren geringere standardisierte Koeffizienten, wobei sich Richtung und Signifikanz nicht ändern. Ein interessanter Mediationseffekt zeigt sich im Hinblick auf den Migrationsstatus. Dessen Effekt verschwindet unter Kontrolle des elterlichen Erziehungsverhaltens. Eine Erklärung lautet, dass Jugendliche mit Migrationshintergrund häufiger elterliche Gewalt erleben (Baier/Pfeiffer 2007), die einen positiven Effekt auf Gewaltviktimsierung hat.

4.1.4 Logistische Regression: Fazit und Ausblick

Die logistische Regression erlaubt die Analyse dichotomer Variablen, wie Viktimisierungsprävalenzen. Im Gegensatz zum linearen Wahrscheinlichkeitsmodell (LPM), das sich aus der OLS-Regression ableitet, spezifiziert die logistische Regression einen nicht linearen Zusammenhang zwischen der vorhergesagten Wahrscheinlichkeit und der unabhängigen Variable, sodass die Interpretation der Ergebnisse schwieriger ist als in der OLS-Regression. Während Logits und Odds Ratios (OR) nur im Hinblick auf Signifikanz und Richtung des Effekts aussagekräftig sind, erlauben Kennzahlen auf Basis vorhergesagter Wahrscheinlichkeiten anschaulichere Aussagen.

Ein Aspekt, der aus Platzgründen nicht ausführlich besprochen werden konnte und vor allem die praktische Durchführung logistischer Regressionen betrifft, ist die *Modelldiagnostik*. Insbesondere die Frage nach dem funktionalen Zusammenhang (Kohler/Kreuter 2008, 283–287) und die Analyse von Residuen und einflussreichen Fällen stehen dabei im Vordergrund (Windzio 2013, 71–72; Long/Freese 2003, 124–127; ausführlich Hosmer/Lemeshow 2000).

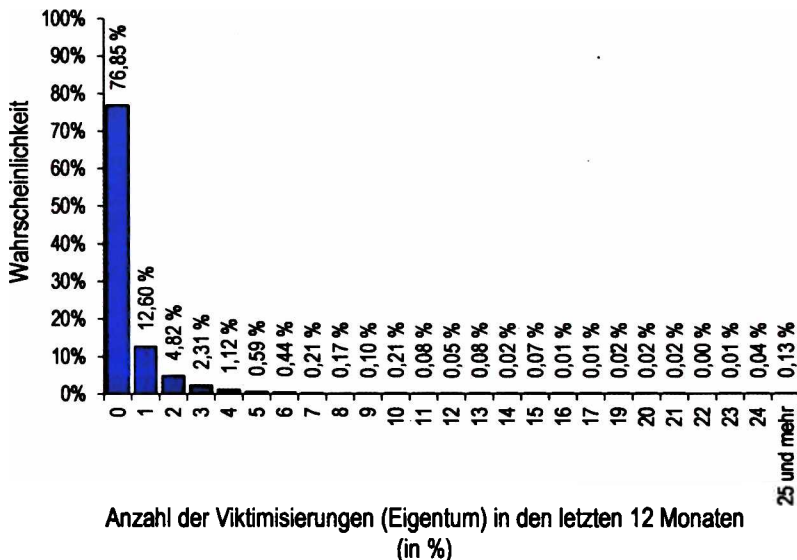
4.2 Zähldaten – Inzidenzen

Im Gegensatz zu Prävalenzen geht es bei Inzidenzen um die Frage, wie häufig ein bestimmtes Ereignis in einem Zeitintervall auftritt. Bezogen auf Opfererfahrungen steht also die Anzahl der Opfererlebnisse in einem bestimmten Zeitraum im Fokus. In der Empirie zeigt sich, dass Inzidenzen oftmals sehr (rechts-)schief verteilt sind, da zahlreiche Personen keine Opfererfahrungen machen. Darüber hinaus haben die meisten Opfer auch nur wenige Viktimisierungen erlebt. Im vorliegenden Beispiel weist die Inzidenz für Eigentumsdelikte eine rechtsschiefe Verteilung auf (*Abbildung 3*). Diese schiefe Verteilung der Inzidenzen kann zu verzerrten Ergebnissen einer OLS-Schätzung führen (Windzio 2013, 193–194).¹⁹

¹⁹ Die Tatsache, dass Inzidenzen diskrete Variablen (nur ganzzahlige Werte) und keine stetig normalverteilten Variablen sind, führt nicht notwendigerweise zu verzerrten Ergebnissen einer OLS-Schätzung (Windzio 2013, 193–194).

Abbildung 3:

Inzidenzen der Viktimisierung durch Eigentumsdelikte (n = 8.411)²⁰



4.2.1 Poisson- und negative Binomialregression

Um dennoch eine Regression schätzen zu können, ist eine Verteilungsfunktion zu verwenden, die der empirischen Verteilung der zu erklärenden Variable entspricht. Im vorliegenden Fall muss die Verteilungsfunktion also diskret (d. h. abzählbar) und asymmetrisch sein. In der Regel wird hierfür die *Poisson-Verteilung* verwendet, die durch einen einzigen, sowohl den Mittelwert als auch die Varianz der Verteilung beschreibenden Parameter μ bestimmt wird. Diese Eigenschaft wird als Equidispersion bezeichnet. Die Poisson-Verteilung gibt die Wahrscheinlichkeit an, dass eine zufällig ausgewählte Person eine bestimmte Zahl von Ereignissen (hier: Viktimisierungen) erlebt hat (Cameron/Trivedi 1998, 1–18; Long/Freese 2003, 245–251; Windzio 2013, 195). Angenommen die empirische Verteilung der Inzidenz von Diebstahl folgt einer Poisson-Verteilung, dann gibt der Mittelwert μ die durchschnittliche Zahl

²⁰ Personen mit 25 und mehr Opfererlebnissen wurden für die Darstellung der Übersichtlichkeit halber zusammengefasst.

der Viktimisierungen in der jeweiligen Stichprobe an. Darüber hinaus ergibt sich aus der Poisson-Verteilung mit dem spezifischen Wert μ die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Person aus der Stichprobe z. B. vier Viktimisierungen erlebt hat. Je höher μ ist, desto mehr verschiebt sich der Modalwert (d. h. der häufigste Wert) der Poisson-Verteilung nach rechts, d. h. die Kategorie mit der höchsten Wahrscheinlichkeit hat eine immer höhere Anzahl an Ereignissen (Windzio 2013, 196–197).

Bei der *Poisson-Regression* (PRM) ergibt sich die Wahrscheinlichkeit einer Beobachtung i (d. h. der Person i), eine bestimmte Zahl von Viktimisierungen erlebt zu haben, aus einer Poisson-Verteilung mit dem Parameter μ_i . Dieser Parameter resultiert aus den Ausprägungen von i der unabhängigen Variablen und den Koeffizienten²¹ und lässt sich auch als Inzidenzrate interpretieren, d. h. als die durchschnittliche Zahl an Viktimisierungen für die jeweilige Beobachtung (Long/Freese 2003, 251–252; Windzio 2013, 196–197).

Ein Problem der PRM ist allerdings, dass die Annahme der Equidispersion, also der Äquivalenz von Mittelwert und Varianz, empirisch nicht immer erfüllt ist (Windzio 2013, 198). Auch im vorliegenden Beispiel zur Inzidenz von Eigentumskriminalität ist dies der Fall. So beträgt die Varianz 7,73, fällt also deutlich größer als der Mittelwert (0,59) aus; man spricht hier von einer Überdispersion, die in der Regel angesichts zahlreicher Fälle, die im Referenzzeitraum kein Ereignis erlebt haben, entsteht. Allerdings ist zu fragen, ob die Abweichung von der Equidispersion auch unter Kontrolle der unabhängigen Variablen weiter fortbesteht (Windzio 2013, 198).

Eine Möglichkeit, um auch in Fällen von Überdispersion Regressions-schätzungen vornehmen zu können, ist die *negative Binomialregression* (NBRM). Zu diesem Zweck wird die Poisson-Verteilung um eine empirisch geschätzte Varianz erweitert. Das PRM ist im NBRM *genestet*,²² sodass beide Modelle miteinander verglichen und mittels statistischen Tests überprüft werden kann, ob tatsächlich eine Abweichung von der Equidispersion vorliegt (Windzio 2013, 198–199). Liegt eine Überdispersion vor, dann sind die Schätzer des PRM ineffizient und die Standardfehler zu klein (Long/Freese 2003, 269).

²¹ Konkret wird der logarithmierte Wert von μ , als Linearkombination der unabhängigen Variablen geschätzt (Windzio 2013, 195)

²² Genestet bedeutet, dass Modell A in Modell B enthalten ist. Dies ist der Fall, wenn Modell B durch das Nullsetzen von Parametern in Modell A überführt werden kann.

4.2.2 Durchführung und Interpretation von Poisson- und negativer Binomialregression

Bei der Analyse von Zähldaten sollte zunächst geprüft werden, ob eine Abweichung von der Equidispersion unter Kontrolle der unabhängigen Variablen vorliegt, d. h. ob der konditionale Mittelwert der konditionalen Varianz entspricht. Das Statistikprogramm Stata berechnet hierzu den Parameter α : Wenn α null ist, dann entspricht die konditionale Varianz dem konditionalen Mittelwert und somit ist das NBRM mit dem PRM identisch (Long/Freese 2003, 268–270; Windzio 2013, 199). In *Tabelle 4* sind für die gleichen Prädiktoren beide Modelle für die Erklärung der Inzidenz von Viktimisierung durch Eigentumskriminalität geschätzt worden. Für α zeigt sich ein Wert von 4,345, der sich signifikant von 0 unterscheidet.²³ Folglich liegt eine Überdispersion vor und es sollte ein NBRM verwendet werden.

Für die Interpretation wurden in *Tabelle 4* die exponierten Koeffizienten angegeben (*Incidence Rate Ratios* = IRR), die sich ähnlich wie die Odds Ratios (OR) interpretieren lassen.²⁴ So rangiert die Viktimisierungsrate (also die durchschnittliche Zahl der erwarteten Ereignisse im gegebenen Zeitraum) von Jungen im Vergleich zu Mädchen um den Faktor 1,16 höher. Analog lassen sich die anderen Dummyvariablen auslegen. Bei den kontinuierlichen Variablen können die IRR als Anstieg der erwarteten Anzahl an Ereignissen interpretiert werden, wenn die unabhängige Variable um eine Einheit steigt. So führt etwa ein Anstieg der Risikosuche um eine Einheit zu einer Erhöhung der erwarteten Zahl an Viktimisierungen um den Faktor 1,186. Ähnlich den OR der logistischen Regression hängt die tatsächliche Rate der Viktimisierungen von den Ausprägungen aller anderen Variablen ab (Long 1997, 224–228; Windzio 2013, 200–201). Die inhaltliche Interpretation der Ergebnisse zu Prädiktoren der Inzidenz von Eigentumsviktimisierung erfolgt unter 4.2.4.

²³ Stata berechnet dazu einen Likelihood-Ratio-Test für die Nullhypothese $\alpha=0$ (StataCorp 2013, 1397).

²⁴ Das PRM wird in Stata mit dem Befehl *poisson* geschätzt, das NBRM mit *nbreg*. Die Option *irr* gibt die Koeffizienten als IRR aus.

Tabelle 4:

Poisson- (PRM) bzw. negatives Binomialmodell (NBRM) für die Inzidenz von Eigentumskriminalität (Incidence Rate Ratios = IRR)

Modell	(1)	(2)
	PRM	NBRM
männlich (1 = ja)	1,295***	1,160*
Migrationshintergrund (1 = ja)	1,043	1,197**
Schulform		
Hauptschule	Referenz	Referenz
Realschule	1,028	0,870
Gymnasium	1,128*	0,957
Risikosuche	1,104***	1,186***
elterliches Monitoring	0,802***	0,833***
elterliche Gewalt		
keine Gewalt	Referenz	Referenz
leichte Gewalt	1,201***	1,308***
schwere Gewalt	1,873***	1,932***
delinquente Freunde		
keine delinquenten Freunde	Referenz	Referenz
1 bis 5 delinquente Freunde	1,912***	1,824***
6 und mehr delinquente Freunde	4,736***	4,607***
Freizeitverhalten		
keine Zeit in Kneipe etc.	Referenz	Referenz
wenig Zeit in Kneipe etc.	0,772***	0,830
moderat Zeit in Kneipe etc.	1,039	1,019
viel Zeit in Kneipe etc.	1,148***	1,156*
Konstante	0,386***	0,337***
Alpha		4,345
McFaddens R ²	0,106	0,040

n = 8.411; *** p < 0,001, ** p < 0,01, * p < 0,05

4.2.3 „Exzess“ von Nullen: *zero-inflated* Modelle

Im vorhergehenden Beispiel wurde deutlich, dass die Angemessenheit des Poisson-Modells (PRM) nicht immer gegeben ist (häufig aufgrund einer sehr hohen Anzahl von Nullen). In manchen Fällen ist die Zahl der Nullen so groß, dass man von einem „Exzess“ der Nullen spricht, der auch durch die negative Binomialverteilung nicht mehr abgebildet werden kann. Ein Exzess von Nullen kann auf soziale Prozesse hindeuten, denen zwei latente Gruppen zugrunde liegen. Diese beiden Gruppen unterscheiden sich signifikant in Bezug auf die abhängige Variable, jedoch gibt es keinen manifesten Faktor, der beide Gruppen identifiziert. Sie sind deshalb unbeobachtet, d. h. latent (Windzio 2013, 201). Bezogen auf die Erklärung von Inzidenzen bedeutet dies zweierlei: Erstens ist zu untersuchen, welche Faktoren die Zugehörigkeit zur Gruppe der Nichtopfer, d. h. Personen die keinerlei Risiko einer Opferwerdung aufweisen, bzw. der potenziellen Opfer, d. h. Personen, die einem Risiko unterliegen, Opfer zu werden, determinieren. Zweitens muss man sich bei den potenziellen Opfern damit befassen, welche Faktoren für die Anzahl der Opfererlebnisse verantwortlich sind. Denkbar ist, dass sich die jeweils relevanten Faktoren hinsichtlich ihres Einflusses unterscheiden. Andere Beispiele aus der Kriminologie wenden diese Idee auf die Erklärung von Gewalttäterschaft an (Windzio/Baier 2009).

Für die Erklärung solcher Phänomene können *zero-inflated* Poisson-Modelle (ZIP) oder *zero-inflated* negative Binomialmodelle (ZINB) verwendet werden. Diese Modelle schätzen den Einfluss der unabhängigen Variablen sowohl auf die Wahrscheinlichkeit, zu einer der beiden Gruppen zu gehören, als auch auf die Anzahl der Ereignisse. Hierfür werden simultan ein binäres Modell zur Erklärung der Viktimisierung (wobei gilt: 0 = Opfer und 1 = Nichtopfer) und ein Modell zur Erklärung der Häufigkeit der Ereignisse berechnet. Aus den vorhergehenden Ausführungen wurde bereits deutlich, dass sich für Ersteres Logitmodelle anbieten, während für die Erklärung der Inzidenz Poisson- oder negative Binomialmodelle infrage kommen (Windzio 2013, 202–203).

Bisher wurden vier Modelle (PRM, NBRM, ZIP und ZINB) zur Erklärung der Inzidenz von Viktimisierung durch Eigentumskriminalität vorgestellt. Alle vier Modelle lassen sich auf die gleichen Daten, im vorliegenden Fall auf die Erklärung von Inzidenzen, anwenden. Daher stellt sich die Frage, welches der vier das angemessene Modell ist, zu deren Beantwortung alle Modelle auf identische Daten angewendet und miteinander verglichen werden:

- Für den Vergleich von PRM und NBRM wurde bereits der Test des Dispersionsparameters α vorgestellt, der im vorliegenden Fall signifikant von 0 verschieden ist, es liegt also keine Equidispersion vor; somit ist das NBRM dem PRM vorzuziehen.
- Um zu überprüfen, ob der Exzess an Nullen so groß ist, dass ein zero-inflated Modell notwendig ist, kann der Vuong-Test²⁵ verwendet werden. Dieser vergleicht das PRM mit dem ZIP bzw. das NBRM mit dem ZINB (Windzio 2013, 204–205). Im vorliegenden Fall wird sowohl für die Poisson-Modelle (nicht dargestellt) als auch für die negativen Binomialmodelle das jeweilige Standardzählmodell zugunsten der zero-inflated Version verworfen. Dies ist aus den signifikanten positiven Werten des Vuong-Tests²⁶ ersichtlich.
- Schließlich kann noch verglichen werden, welches der beiden zero-inflated Modelle ZIP und ZINB angemessener ist. Da beide Modelle genestet sind, kann dies mittels eines Likelihood-Ratio-Tests überprüft werden. Analog zum Vergleich von PRM und NBRM wird getestet, ob sich der Überdispersionsparameter α signifikant von 0 unterscheidet (Long/Freese 2003, 285). Im vorliegenden Fall verwirft der Likelihood-Ratio-Test²⁷ die Nullhypothese $H_0: \alpha = 0$ mit $p < 0,001$, sodass das ZINB dem ZIP vorgezogen wird.²⁸

4.2.4 Modellschätzung und Interpretation

Tabelle 5 gibt die Ergebnisse des zero-inflated negativen Binomialmodells (ZINB) für die Inzidenz von Eigentumsviktimisierung wieder. Hierbei werden für jede Variable zwei Koeffizienten ausgegeben. Der erste Koeffizient veranschaulicht den Einfluss der Variable auf die Wahrscheinlichkeit, zur Gruppe der Nichtopfer ($Y=0$) zu gehören. Die Koeffizienten sind hier als Odds Ratio (OR) dargestellt und werden analog zur logistischen Regression interpretiert. Im Ergebnis zeigt sich, dass Jungen eine höhere Wahrscheinlichkeit haben, zur Gruppe der *Nichtopfer* zu gehören (OR von 2,025). Gleichzei-

²⁵ Für eine detailliertere Darstellung des Tests siehe Windzio 2013, 204–205 und Long 1997, 248–249. In Stata kann der Test als Option (*vuong*) bei der Schätzung eines ZIP- bzw. ZINB-Modells angefordert werden.

²⁶ Die Teststatistik für den Vergleich von NBRM und ZINB ist in *Tabelle 5* angegeben.

²⁷ Siehe Long und Freese 2003, 285 für Details und Teststatistik.

²⁸ Daneben existiert auch eine grafische Methode, die die auf Basis des jeweiligen Modells vorhergesagten Wahrscheinlichkeiten mit den beobachteten Wahrscheinlichkeiten vergleicht (hierzu Long 1997, 247–249; Long/Freese 2003, 283–284; Windzio 2013, 205–206).

tig beeinflusst das Geschlecht auch die Inzidenz signifikant: Jungen haben eine höhere Inzidenzrate als Mädchen (IRR von 1,265). Eine Erklärung könnte lauten, dass Jungen eher dazu neigen, Eigentumskriminalität nicht als solche wahrzunehmen und dementsprechend nicht im Fragebogen angeben. Diejenigen, die aber derartige Erlebnisse als Kriminalität interpretieren und angeben, erleben Derartiges tatsächlich häufiger. Andere Prädiktoren zeigen demgegenüber konsistente Effekte. So weisen Jugendliche, die delinquente Freunde haben, eine höhere Wahrscheinlichkeit auf, in der Gruppe der Opfer zu sein, und gleichzeitig erhöht die Präsenz delinquenter Freunde die Inzidenzrate. Risikosuche hat nur einen Effekt auf die Gruppenzugehörigkeit von Opfern und Nichtopfern, nicht jedoch auf die Häufigkeit der Opferwerdung. Elterliches Monitoring hat keinen Einfluss auf die Gruppenzugehörigkeit, reduziert aber die Zahl der Viktimisierungen signifikant. Elterliche Gewalt erhöht die Wahrscheinlichkeit, zur Gruppe der Opfer zu gehören (nur leichte Gewalt signifikant), und die Zahl der Opferwerdungen (nur schwere Gewalt signifikant).

Abschließend sei auf weitere Möglichkeiten von Interpretation und Darstellung der Ergebnisse von Zähldatenmodellen verwiesen. Analog zum logistischen Modell lassen sich verschiedene Kennziffern des Zusammenhangs auf Basis vorhergesagter Wahrscheinlichkeiten berechnen und grafisch darstellen (Long/Freese 2003).

Tabelle 5:

Zero-inflated negatives Binomialmodell zur Erklärung der Inzidenz von Eigentumsviktimisierung

	ZINB	
	OR (Y = 0)	IRR
männlich (1 = ja)	2,025*	1,265***
Migrationshintergrund (1 = ja)	0,753	1,140
Schulform		
Hauptschule	Referenz	Referenz
Realschule	1,232	0,927
Gymnasium	1,166	0,998
Risikosuche	0,450**	1,080
elterliches Monitoring	1,002	0,835***

	ZINB	
	OR (Y = 0)	IRR
elterliche Gewalt		
keine Gewalt	Referenz	Referenz
leichte Gewalt	0,165*	1,091
schwere Gewalt	0,580	1,764***
delinquente Freunde		
keine delinquenten Freunde	Referenz	Referenz
1 bis 5 delinquente Freunde	0,686	1,711***
6 und mehr delinquente Freunde	0,982	4,365***
Freizeitverhalten		
keine Zeit in Kneipe etc.	Referenz	Referenz
wenig Zeit in Kneipe etc.	0,105	0,645**
moderat Zeit in Kneipe etc.	0,274*	0,839
viel Zeit in Kneipe etc.	0,464	1,002
Konstante	1,776	0,565*
Vuong		2,698**
Alpha		3,584
McFaddens R ²		0,043

n = 8.411; *** p < 0,001, ** p < 0,01, * p < 0,05

4.3 Mehrebenenmodelle

In den vorherigen Abschnitten bestand die Problematik darin, dass Skalierung bzw. Verteilung der abhängigen Variable eine Anwendung der OLS-Regression ausschließt. In diesem Abschnitt steht demgegenüber die Struktur der Daten selbst im Fokus. Konkret geht es darum, dass die Beobachtungen in Kontexten enthalten sind; man spricht in diesem Fall auch von geklumpten, geclusterten oder hierarchischen Daten. Als Beispiele für hierarchische Datenstrukturen lassen sich etwa Schülerinnen und Schüler in Schulklassen, Bewohnerinnen und Bewohner einer Nachbarschaft oder auch Straftäterinnen und Straftäter in Gefängnissen anführen. Dabei wird die kleinste Ebene (z. B. Schülerinnen und Schüler) als Level-1 bezeichnet und die Kontextebene als Level-2 (z. B. Schulklasse). Dieses Design lässt sich auch erweitern. So wäre

bspw. die zusätzlich eingefügte Ebene der Schule Level-3. In den im Folgenden diskutierten Modellen liegt die abhängige Variable immer auf der untersten Ebene (Level-1).

Weisen Daten eine solche hierarchische Struktur auf, dann werden alle Beobachtungen dieses Kontexts durch identische Umweltbedingungen beeinflusst. In der Folge – vorausgesetzt diese Umweltbedingungen beeinflussen die abhängige Variable – sind sich die Beobachtungen eines Kontexts ähnlicher als Beobachtungen aus verschiedenen Kontexten. Dies führt aber zur Verletzung der Annahme statistischer Unabhängigkeit der Beobachtungen, die die meisten statistischen Verfahren voraussetzen (Windzio 2008, 113–114).

Eine Maßzahl zur Beschreibung der Ähnlichkeit von Einheiten desselben Kontexts ist der *Intraclass Correlation Coefficient* (ICC), definiert als der Quotient aus der Varianz zwischen den Kontexten und der Gesamtvarianz. Die Maßzahl lässt sich interpretieren als der Anteil der Varianz der abhängigen Variable, der durch die Gruppenstruktur erklärt wird. Je höher der ICC ist, desto ähnlicher sind sich Beobachtungen aus demselben Kontext (für die Berechnung siehe Snijders/Bosker 2012, 17–23).

Die hierarchische Datenstruktur kann dabei einerseits ein unerwünschter Nebeneffekt des Stichprobendesigns sein. So wird aus praktischen Erwägungen und aus Kostengründen oft ein mehrstufiges Stichprobendesign verwendet, bei dem zunächst die Level-2-Einheiten gezogen werden und aus diesen dann die zu untersuchenden Level-1-Einheiten. Andererseits kann eine hierarchische Datenstruktur aber auch helfen, interessante soziale Phänomene insofern aufzudecken, als Eigenschaften des Kontexts direkt oder indirekt Eigenschaften der Individuen beeinflussen (Snijders/Bosker 2012, 6–9; Windzio 2008).

Allgemein betrachtet lassen sich mittels eines Mehrebenen-Designs drei Arten von Hypothesen überprüfen. So können zwei Level-1-Variablen in Beziehung gesetzt werden (Mikrohypothese), um z. B. den Einfluss der elterlichen Erziehung auf das Risiko der Opferwerdung zu untersuchen. In diesem Fall wäre die Multilevelstruktur eher als „Nuisance“ (Störfaktor)²⁹ zu betrachten, den es bei der Analyse zu berücksichtigen gilt. Daneben können aber auch substantielle Hypothesen zum Einfluss des Kontexts geprüft werden. Makro-Mikro-Hypothesen untersuchen den Effekt von Level-2-Variablen (z. B. Klassenklima) auf Level-1-Variablen (Viktimisierungsrisiko). Cross-Level-Interaktionen schließlich befassen sich mit dem Effekt von Level-2-Variablen auf

²⁹ Ist man nicht an der Aufdeckung von Kontexteffekten interessiert, können statt Mehrebenenmodellen auch normale Regressionsverfahren mit Standardfehlern geschätzt werden, die gegenüber Klumpenstichproben robust sind („geclusterte“ Standardfehler).

die Beziehung zwischen zwei Level-1-Variablen. So ist es denkbar, dass das Klassenklima den Zusammenhang zwischen der elterlichen Erziehung der Schülerinnen und Schüler und dem Viktimisierungsrisiko beeinflusst, da insbesondere in einem positiven Umfeld Defizite von benachteiligten Schülern ausglich werden können (Snijders/Bosker 2012, 9–12).

4.3.1 Das Mehrebenenmodell

Die nachfolgende Darstellung versucht möglichst ohne mathematische Formeln auszukommen und lehnt sich an Windzio (2008) an. Es wird im Folgenden von Personen in Kontexten die Rede sein, die Ausführungen lassen sich aber allgemein auf hierarchische Daten übertragen. Ausgangspunkt der Darstellung der Mehrebenenanalyse ist das oben vorgestellte lineare Regressionsmodell. Der Wert der abhängigen Variablen y_i für Person i hängt von einem Intercept β_0 , dem Koeffizienten β_1 , der Ausprägung der unabhängigen (Level-1-)Variable x_{1i} für das Individuum i und dem Fehlerterm ϵ_i ab.

$$y_i = \beta_0 + \beta_1 x_{1i} + \epsilon_i \quad (3)$$

Im Fall von Personen, die in Kontexten geclustert sind, kann Gleichung (3) um einen eigenen Intercept β_{0j} für jeden Kontext erweitert werden (4). Die abhängige Variable y_{ij} gibt somit den Wert der i ten Person in Kontext j an, der neben dem Fehlerterm ϵ_{ij} von den Werten von x_{1ij} , dem Koeffizienten β_1 und dem Intercept β_{0j} des Kontexts j , dem die jeweilige Person angehört, abhängt.

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \epsilon_{ij} \quad (4)$$

Der spezifische Intercept β_{0j} für einen Kontext j (z. B. für eine bestimmte Klasse) ergibt sich aus dem Mittelwert der Intercepts über alle Kontexte β_{00} und den Abweichungen μ_{0j} der einzelnen Kontexte von diesem Mittelwert. Ist die Abweichung μ_{0j} negativ, dann weisen die Mitglieder dieses Kontexts im Durchschnitt geringere Werte für y auf als der Durchschnitt aller Beobachtungen (kontrolliert um den Effekt der unabhängigen Variablen x). Die kontextspezifischen Abweichungen μ_{0j} entstammen dabei einer Zufallsverteilung, die aus den Daten geschätzt wird. Die Varianz dieser Verteilung gibt daher an, wie stark sich die Mittelwerte der Kontexte voneinander unterscheiden. Wäre die Varianz 0, würden wir keine Unterschiede zwischen den Kontexten beobachten. Das Modell in Gleichung (4) wird als *Random-Intercept-Modell* bezeichnet.

Die Logik der Varianz des Intercepts lässt sich auch auf die Koeffizienten der unabhängigen Variablen übertragen. Hierzu erweitert man die zu schätzende Gleichung um einen eigenen Koeffizienten (Slope) β_{1j} für jeden Kontext. Das

bedeutet im Fall der linearen Regression, dass die Steigungen der Regressionsgeraden zwischen den Kontexten variieren (*Random-Slope-Modell*).

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \varepsilon_{ij} \tag{5}$$

Analog zum Random-Intercept-Modell ergibt sich der Koeffizient β_{1j} für den Kontext j aus dem Mittelwert β_{10} über alle kontextspezifischen Slopes und der Abweichung μ_{1j} des Kontexts j von diesem Mittelwert. μ_{1j} entstammt ebenfalls einer Zufallsverteilung, die aus den Daten geschätzt wird. Je größer deren Varianz ist, desto stärker unterscheiden sich die Slopes zwischen den Kontexten. Variieren die Slopes signifikant (d. h. die Varianz von μ_1 ist größer 0), dann bedeutet dies, dass sich die Zusammenhänge zwischen der unabhängigen und abhängigen Variable zwischen den Kontexten unterscheiden.

Die Mehrebenenanalyse ist dementsprechend in der Lage, Unterschiede sowohl in den Intercepts als auch in den Slopes zu modellieren. Es ist jedoch in beiden Fällen eine empirische Frage, ob beide signifikant zwischen den Kontexten variieren, d. h. ob sich die Varianzen von μ_0 und μ_1 signifikant von null unterscheiden.

4.3.2 Die hierarchische Datenstruktur als erklärender Faktor

Der Nutzen der Mehrebenenanalyse beschränkt sich gleichwohl nicht darauf festzustellen, ob sich die Intercepts oder die Slopes zwischen Kontexten unterscheiden. Die Analyse kann zugleich zur Erklärung sozialer Prozesse genutzt werden. Zeigen Analysen, dass die Intercepts signifikant zwischen den Modellen variieren, kann versucht werden, diese Unterschiede durch Eigenschaften der Kontexte zu erklären. Im eingangs genannten Beispiel (Makro-Mikro-Hypothese) könnte etwa das Klassenklima das Viktimisierungsrisiko beeinflussen. Um dies zu überprüfen, kann eine Gruppenvariable z (die den gleichen Wert für alle Beobachtungen eines Kontexts hat) in das Modell aufgenommen werden. Diese Variable kann die Varianz der Abweichungen der Kontexte μ_0 reduzieren, d. h. erklären. Im Fall eines Random-Slope-Modells würde sich Gleichung (5) erweitern zu:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{01}z_j + \varepsilon_{ij} \quad \text{mit } \beta_{0j} = \beta_{00} + \mu_{0j} \tag{6}$$

Variieren in einem Modell die Slopes einer (oder mehrerer) Level-1-Variablen signifikant (Random-Slope-Modell), kann auch versucht werden, diese zu erklären. Hierfür wird ein Interaktionseffekt zwischen der Level-1-Variablen mit dem Random Slope und einer Level-2-Variablen spezifiziert. Dies bedeutet, dass die Varianz des Koeffizienten der Level-1-Variable durch die Level-2-Variable erklärt wird (ausführlich Snijders/Bosker 2012; Windzio 2008).

4.3.3 Durchführung und Interpretation einer logistischen Mehrebenenanalyse

Im Folgenden soll die Durchführung und Interpretation einer Mehrebenenanalyse am Beispiel der Erklärung von Viktimisierung durch Schulgewalt verdeutlicht werden. Die Betrachtung des Schulkontexts erlaubt dabei, Merkmale der Klasse als substantielle Einflussfaktoren zu betrachten. Für die Erklärung der Prävalenz von Viktimisierung durch Schulgewalt wird ein logistisches Mehrebenenmodell³⁰ verwendet. Durch die nicht lineare Linkfunktion kommen etwas andere Formeln zur Anwendung als eben für das lineare Modell dargestellt, die Mehrebenenlogik bleibt aber bestehen. Zusätzlich sind die Schwierigkeiten bei der Interpretation der Koeffizienten zu beachten, auf die in Unterkapitel 4.1 hingewiesen wurde.

Tabelle 6 zeigt die Ergebnisse der Modellschätzung.³¹ In einem ersten Schritt wird das sogenannte leere Modell, das nur Random Intercepts und keine erklärenden Variablen enthält, geschätzt, um den Grad der Ähnlichkeit zwischen Individuen derselben Klasse zu bestimmen (Modell 0). Im Ergebnis zeigt sich ein Intraklassenkorrelationskoeffizient (ICC) von 0,074. Dementsprechend gehen 7,4 % der Varianz der Viktimisierung durch Schulgewalt auf die Klassen zurück. Die Varianz der Intercepts ist signifikant, wie ein Likelihood-Ratio-Test der Nullhypothese, dass die Varianz der Random Intercepts gleich null ist, zeigt.³²

In Modell 1 wird der Einfluss der Level-1-Variablen auf das Risiko, Schulgewalt zu erleben, geschätzt. Die metrischen Variablen wurden am Gesamtmittelwert (*Grand Mean*) zentriert. Dies ist bei Mehrebenenmodellen aus zwei Gründen sinnvoll. Zum einen erhält in diesem Fall die Konstante einen sinnvollen Wert und zum anderen hängt die Varianz der Random Slopes davon ab, ob die Variablen zentriert sind oder nicht (Windzio 2008, 126–128).³³ In diesen Modellen wurde auf die Indikatoren für Alkoholkonsum und den Besuch von „Zeit pro Woche in Kneipe, Disco, Kino etc.“ verzichtet, da diese im Sinne des Routine-Activity-Ansatzes eher Prädiktoren für Viktimisierung

³⁰ Ausführlicher hierzu Windzio 2008, 129–134 sowie Snijders und Bosker 2012, 290–309.

³¹ Alle Modelle wurden in Stata mit dem Befehl *xmelogit* geschätzt. Es wird ein *Unit-specific*-Modell geschätzt. Für eine Diskussion der Unterschiede zwischen *Unit-specific*- und *Population-Average*-Modell siehe Raudenbush u. a. 2004, 109–11 und Neuhaus u. a. 1991.

³² Dieser Test wird im Anschluss an die Schätzung von Stata standardmäßig ausgegeben.

³³ Eine andere Form der Zentrierung, die im Rahmen von Mehrebenenanalysen anzutreffen ist, ist die *Group-Mean*-Zentrierung, die das Testen sogenannter Frotschteichypothesen erlaubt Windzio 2008, 127–128.

außerhalb des Schulkontexts darstellen. In der Tabelle sind die OR dargestellt, die sich analog zur einfachen logistischen Regression interpretieren lassen. Es zeigt sich, dass Jungen ein mehr als dreimal höheres Risiko haben, Opfer von Schulgewalt zu werden, als Mädchen. Migrantinnen und Migranten haben demgegenüber ein geringeres Risiko als Einheimische. Der besuchte Schultyp hat keinen signifikanten Einfluss auf das Viktimisierungsrisiko, wengleich eine höhere elterliche Supervision der Jugendlichen mit einem geringeren Risiko für Gewaltviktimsierung einhergeht. Elterliche Gewalt und delinquente Freundinnen und Freunde erhöhen demgegenüber das Risiko, Opfer von Schulgewalt zu werden. Der ICC sinkt durch Hinzunahme der Level-1-Variablen leicht auf 0,072 ab, sodass ein Teil der Varianz zwischen den Klassen durch eine unterschiedliche Zusammensetzung der Klassen im Hinblick auf diese Level-1 Variablen erklärt wird (Kompositionseffekt).

Modell 2 berücksichtigt zusätzlich zu den Level-1-Variablen den Einfluss von Kontextmerkmalen. Es werden das Klassenniveau an Desorganisation und Kohäsion in das Modell aufgenommen. Während eine höhere schulische Desorganisation mit einem erhöhten Viktimisierungsrisiko einhergeht, ist der Einfluss der Kohäsion nicht signifikant. Der ICC sinkt gegenüber Modell 1 leicht ab, d. h., ein Teil der Varianz zwischen den Klassen kann durch die Kontextmerkmale erklärt werden. Die Varianz des Intercepts $\text{Var}(\mu_{0j})$ ist in Modell 2 signifikant größer als null.

Im nächsten Schritt (Modell 3) wird neben dem Random Intercept ein Random Slope für den Einfluss des elterlichen Monitorings geschätzt, d. h., es wird überprüft, ob der Effekt des Monitorings der Eltern auf das Viktimisierungsrisiko zwischen den Klassen variiert. Die Idee dahinter ist, dass der Effekt elterlichen Monitorings von den Eigenschaften des Klassenkontexts abhängt. So konnte in der bisherigen Forschung (Hanslmaier 2014) gezeigt werden, dass der Einfluss familialen Sozialkapitals auf Gewalttäterschaft vom schulischen Sozialkapital abhängt. Allerdings ist zu überprüfen, ob der Zusammenhang zwischen elterlichem Monitoring und dem Viktimisierungsrisiko (also die Slopes) tatsächlich über die Klassen signifikant variiert. Wenn die Slopes signifikant variieren, dann bedeutet dies, dass die Varianz der Abweichungen der Slopes für elterliches Monitoring von der mittleren Steigung, $\text{Var}(\mu_{1j})$, signifikant ist. Dies wird mit einem Deviance-Test überprüft, der das Modell ohne Random Slope mit dem Modell mit Random Slope vergleicht.³⁴ Der Test zeigt ein nicht signifikantes Ergebnis. Das bedeutet, dass der Zusam-

³⁴ Der in diesem Fall von Stata ausgegebene Test ist konservativ. Deshalb empfiehlt es sich, den in Snijders und Bosker 2012, 98–99 beschriebenen Test, der auf der Differenz der von Stata berechneten *Deviances* der Modelle beruht, zu verwenden. Die kritischen Werte ergeben sich dabei aus einer speziellen Tabelle.

menhang zwischen elterlicher Supervision und Viktimisierung durch Schulgewalt nicht signifikant zwischen den Schulklassen variiert. In Modell 3 steigt auch der ICC gegenüber Modell 2 an.

Tabelle 6:

Mehrebenenmodell zur Erklärung der Prävalenz der Viktimisierung durch Schulgewalt (OR)³⁵

	(0)	(1)	(2)	(3)	(4)
<i>Level-1-Variablen</i>					
männlich (1 = ja)		3,073***	3,085***	3,100***	3,079***
Migrationshintergrund (1 = ja)		0,735***	0,724***	0,724***	0,724***
Schulform					
Hauptschule					
Realschule		1,135	1,250	1,276	1,240
Gymnasium		0,956	1,232	1,253	1,228
Risikosuche		1,096*	1,096*	1,098*	1,097*
elterliches Monitoring		0,852***	0,854***	0,872**	0,841**
elterliche Gewalt					
keine Gewalt					
leichte Gewalt		1,468***	1,468***	1,470***	1,465***
schwere Gewalt		2,060***	2,036***	2,037***	2,029***
delinquente Freunde					
keine delinquenten Freunde					
1 bis 5 delinquente Freunde		1,590***	1,576***	1,574***	1,572***
6 und mehr delinquente Freunde		1,665***	1,625***	1,636***	1,626***
<i>Level-2-Variablen</i>					
schulische Desorganisation			1,626**	1,657***	1,666***
schulische Kohäsion			0,972	0,976	0,972
elterliches Monitoring x schulische Desorganisation					1,226*

³⁵ Die ICCs in der Tabelle wurden mit dem Paket *xtrho* von Lars E. Kroll berechnet. Für die Berechnung des R² wurde das Paket *r2_mz* von Dirk Enzmann verwendet. Alternativ können beide Maßzahlen auch von Hand auf Basis der entsprechenden Formeln in Snijders und Bosker 2012 berechnet werden.

	(0)	(1)	(2)	(3)	(4)
Var(μ_{0j})	0,263	0,256	0,237	0,232	0,237
Var(μ_{1j})				0,039	
Cov (μ_{0j} ; μ_{1j})				-0,036	
ICC	0,074	0,072	0,067	0,076	0,067
McKelvey & Zavoina's R ²	-	0,136	0,141	0,140	0,143

n = 8.411 Schüler aus 454 Klassen. *** p < 0,001, ** p < 0,01, * p < 0,05, + p < 0,10; alle metrischen Variablen sind Grand-Mean-zentriert

Eingangs wurde argumentiert, dass ein Cross-Level-Interaktionseffekt spezifiziert wird, um die Varianz des Slopes einer Level-1-Variable zu erklären. Allerdings ist es mitunter auch ratsam, von dieser Strategie abzuweichen, nämlich dann, wenn ein Cross-Level-Interaktionseffekt aus theoretischen Gründen zu erwarten ist. Dies ist insbesondere deshalb sinnvoll, da die statistische Power zur Aufdeckung einer Cross-Level-Interaktion höher als die den korrespondierenden Random Slope aufdeckende ist (wenn tatsächlich ein solcher Interaktionseffekt existiert) (Snijders/Bosker 2012, 106).

Modell 4 schätzt dementsprechend einen Cross-Level-Interaktionseffekt zwischen elterlichem Monitoring und dem Level schulischer Desorganisation. Der Koeffizient der Cross-Level-Interaktion ist marginal signifikant (p < 0,10) und hat ein OR > 1. Für die Interpretation bedeutet dies, dass in Kontexten mit einem mittleren Niveau³⁶ an Desorganisation ein Anstieg der elterlichen Supervision zu einem geringeren Risiko der Viktimisierung führt. In Kontexten mit einer geringeren Desorganisation führt der Anstieg elterlichen Monitorings zu einer größeren Reduzierung des Viktimisierungsrisikos. Demgegenüber ist der Effekt elterlichen Monitorings in Kontexten mit hoher Desorganisation geringer. Allerdings sollte der Interaktionseffekt aufgrund der relativ hohen Irrtumswahrscheinlichkeit eher zurückhaltend interpretiert werden. Zudem ist zu berücksichtigen, dass in logistischen Modellen bereits implizit Interaktionseffekte enthalten sind, da der Einfluss einer Variablen von den Ausprägungen der anderen Variablen abhängt. Dies erschwert den Nachweis von Interaktionseffekten (Best/Wolf 2010, 840–842; ausführlich Ai/Norton 2003). Auch ein Blick auf die Modellfitparameter (McKelvey & Zavoina's R²) zeigt kaum Verbesserungen zwischen Modell 2 und Modell 4. Der ICC bleibt gleich, d. h., Unterschiede zwischen den Kontexten können

³⁶ Da alle metrischen Variablen am Gesamtmittelwert (Grand-Mean) zentriert wurden, bedeutet der Wert Null bei dieser Variable, dass ein mittleres Niveau an Desorganisation vorliegt.

auch nicht besser erklärt werden, wenn man den Interaktionsterm mit in das Modell aufnimmt. Auch das McKelvey & Zavoina's R^2 steigt nur gering an.

Für die Beispielanalyse standen 8.411 Schülerinnen und Schüler aus 454 Klassen zur Verfügung. Generell gilt im Rahmen der Mehrebenenanalyse, dass die Anzahl der Level-1-Einheiten für Schätzungen von Level-1-Effekten am wichtigsten ist und die Zahl der Level-2-Einheiten für die Schätzung von Level-2-Effekten. Die Größe der Cluster spielt keine besondere Rolle. Die Schätzung von Random Slopes hingegen ist von der Größe der Cluster abhängig (Snijders 2005). Weiterführende Diskussionen zu statistischer Power, d. h. zum Aufdecken eines tatsächlich vorhandenen Effekts, in Abhängigkeit des Studiendesigns finden sich bei Snijders und Bosker (2012).

5 Zusammenfassung

- Daten aus Opferbefragungen weisen bestimmte Eigenschaften auf, die die Anwendung einfacher linearer Regressionen (OLS-Regressionen) verhindern, da bestimmte statistische Annahmen verletzt werden. Dies betrifft die Skalierung und Verteilung der abhängigen Variable. So stellen Prävalenzen von Opfererfahrungen dichotome Merkmale dar und erfordern daher spezielle Regressionsverfahren, die diesem Umstand Rechnung tragen. Bei der Analyse von Inzidenzen von Opfererfahrungen tritt das Problem auf, dass diese in der Regel sehr rechtsschief verteilt sind. Darüber hinaus kann als Folge der Stichprobenziehung eine hierarchische Datenstruktur oder spezielles Interesse an Kontexteinflüssen vorliegen.
- Die *logistische Regression* löst das Problem der multivariaten Analyse dichotomer Merkmale. Hierfür wird eine Linkfunktion verwendet, die dazu führt, dass die Beziehung zwischen den unabhängigen Variablen und der Wahrscheinlichkeit, eine Viktimisierung zu erleben, nicht mehr linear ist. Dies erschwert die Interpretation der Koeffizienten im Vergleich zur linearen Regression.
- *Zählmodellen* kommen bei der Analyse von Inzidenzen zur Anwendung. Die Poisson- und die Negativbinomialregression verwenden hierfür eine Verteilungsfunktion, die der schiefen und diskreten Verteilung der Daten angemessen ist. Wenn die Zahl der Nullen sehr groß ist, d. h. es gibt viele Personen, die überhaupt nicht Opfer geworden sind, dann kommen zero-inflated Modelle zur Anwendung. Bei diesen Verfahren werden simultan zwei Modelle geschätzt, die einerseits den Einfluss der unabhängigen Variablen auf die Wahrscheinlichkeit, zur Gruppe der Nichtopfer zu

gehören, und andererseits den Einfluss dieser Variablen auf die Häufigkeit der Opfererfahrung angeben (Windzio 2013).

- Zudem weisen Daten aus größeren Surveys oftmals eine *hierarchische Struktur* auf, d. h., die Beobachtungen auf der Individualebene sind in bestimmten Kontexten (z. B. Schulen, Nachbarschaften) geschachtelt. Dies führt zu einer Verletzung der Annahme der statistischen Unabhängigkeit von zwei Beobachtungen.
- *Mehrebenenmodelle* dienen der Analyse geclusterter Daten. Ihre Struktur (z. B. Schülerinnen und Schüler in Klassen) kann dabei entweder als Störfaktor betrachtet werden, den es zu kontrollieren gilt, um unverzerrte Ergebnisse zu erhalten, oder dazu genutzt werden, soziale Prozesse, wie etwa den Einfluss von Kontexten, aufzudecken (Windzio 2008).

6 Weiterführende Literatur

Der Aufsatz von Best und Wolf (2010) liefert eine kurze, aber umfassende Darstellung der logistischen Regression. Auch das entsprechende Kapitel bei Windzio (2013) ist zu empfehlen. Eine ausführlichere Diskussion findet sich bei Long (1997). Windzio und Long behandeln zudem in ihren Büchern die hier besprochenen Zähldatenmodelle. Im Hinblick auf die praktische Durchführung von logistischen und Zähldatenanalysen mit Stata bietet sich insbesondere das Buch von Long und Freese (2003) an, das Beispiele aufführt und den Möglichkeiten der Ergebnisdarstellung ausführlich Raum gibt. Kohler und Kreuter (2008) liefern ebenfalls eine gute Einführung in die Durchführung logistischer Regressionen mit Stata.

Einen kurzen, sehr anschaulichen Überblick über die Logik der Mehrebenenanalyse liefert Windzio (2008). Ein ausführlicheres „Standardwerk“ ist das Buch *Multilevel Analysis* von Snijders und Bosker (2012). Dort werden auch die Voraussetzungen des Modells, Strategien zur Modellbildung und verschiedene Regressionsverfahren (OLS, Logit, Zähldatenmodelle) erläutert. Das zweibändige Werk *Multilevel and Longitudinal Modeling Using Stata* von Rabe-Hesketh und Skrondal (2012) bietet eine Einführung für Stata mit zahlreichen Übungsbeispielen und konkreten Hinweisen für die Durchführung der Analysen.

8 Literatur

- Ai, Chunrong; Norton, Edward C. (2003): Interaction Terms in Logit and Probit Models. *Economics Letters*, 80, S. 123–129.
- Baier, Dirk (2015): Sicherheit und Kriminalität in Niedersachsen, Ergebnisse einer Repräsentativbefragung, KFN-Forschungsbericht Nr. 127. Hannover: KFN.
- Baier, Dirk; Pfeiffer, Christian (2007): Gewalttätigkeit bei deutschen und nichtdeutschen Jugendlichen – Befunde der Schülerbefragung 2005 und Folgerungen für die Prävention, KFN-Forschungsbericht Nr. 100. Hannover: KFN.
- Baier, Dirk; Pfeiffer, Christian; Simonson, Julia und Rabold, Susann (2009): Jugendliche in Deutschland als Opfer und Täter von Gewalt. Erster Forschungsbericht zum gemeinsamen Forschungsprojekt des Bundesministeriums des Innern und des KFN. Hannover: KFN.
- Baier, Dirk; Prätör, Susann (im Druck): Adolescents as Victims of Violence. In: *Representative Studies on Victimization. Research Findings from Germany*.
- Bartus, Tamás (2005): Estimation of marginal effects using `margins`. In: *Stata Journal*, 5, S. 309–329.
- Best, Henning; Wolf, Christof (2010): Logistische Regression. In: Wolf, Christof; Best, Henning (Hg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag, S. 827–854.
- Cameron, Colin A.; Trivedi, Pravin K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.
- Cohen, Lawrence E.; Felson, Marcus (1979): Social Change and Crime Rate Trends: A Routine Activity Approach. In: *American Sociological Review*, 44, S. 588–608.
- Enzmann, Dirk (2010): Chapter 4. Germany. In: Junger-Tas, Josine; Marshall, Ineke Haen; Enzmann, Dirk; Killias, Martin; Steketee, M. und Gruszczynska, Beata (Hg.): *Juvenile Delinquency in Europe and Beyond. Results of the Second International Self-Report Delinquency Study*. New York, NY: Springer New York, S. 47–64.
- Gautschi, Thomas (2010): Maximum-Likelihood Schätztheorie. In: Wolf, Christof; Best, Henning (Hg.): *Handbuch der sozialwissenschaftlichen Datenanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 205–235.
- Gottfredson, Michael R.; Hirschi, Travis (1990): *A General Theory of Crime*. Stanford California: Stanford University Press.
- Gottfredson, Michael R.; Hirschi, Travis (2001): Self-Control Theory. In: Paternoster, Raymond; Bachman, Ronet (Hg.): *Explaining Criminals and Crime*. Los Angeles, CA: Roxbury Publishing Company, S. 81–96.

- Gruszczynska, Beata; Lucia, Sonia und Killias, Martin (2012): Chapter 4. Juvenile Victimization from an International Perspective. In: Junger-Tas, Josine; Marshall, Ineke Haen; Enzmann, Dirk; Killias, Martin; Steketee, Majona und Gruszczynska, Beata (Hg.): *The Many Faces of Youth Crime*. New York, NY: Springer New York, S. 95–116.
- Hanslmaier, Michael (2014): Soziales Kapital und Jugendgewalt. Die Wechselwirkungen von Schule und Familie. In: *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 34, S. 314–330.
- Hosmer, David W.; Lemeshow, Stanley (2000): *Applied Logistic Regression*, 2. Auflage. New York: John Wiley.
- Karolson, Kristian B.; Holm, Anders und Breen, Richard (2012): Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method. In: *Sociological Methodology*, 42, S. 286–313.
- Kohler, Ulrich; Kreuter, Frauke (2008): *Datenanalyse mit Stata*, 3. Auflage. München: Oldenbourg.
- Lee, Eun S.; Forthofer, Ronald N. (2006): *Analyzing Complex Survey Data*, 2. Auflage. Sage: Thousand Oaks.
- Long, J. Scott (1997): *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.
- Long, J. Scott; Freese, Jeremy (2003): *Regression Models for Categorical Dependent Variables Using Stata*. College Station, Texas: Stata Corporation.
- Mood, Carina (2009): Logistic regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. In: *European Sociological Review*, 26, S. 67–82.
- Neuhaus, J., Kalbfleisch, J. und Hauck, W. (1991): Comparison of Cluster-Specific Approaches for Population-Averaged Data Analyzing Correlated. *International Statistical Review*, 59, 1, S. 25–35.
- Oberwittler, Dietrich (2003): Die Messung und Qualitätskontrolle kontextbezogener Befragungsdaten mithilfe der Mehrebenenanalyse – am Beispiel des Sozialkapitals von Stadtvierteln. In: *ZA-Informationen*, 53, S. 11–41.
- Paternoster, Raymond; Bachman, Ronet (2001a): Classical and Neue Classical Schools of Criminology. In: Paternoster, Raymond; Bachman, Ronet (Hg.): *Explaining Criminals and Crime*. Los Angeles, CA: Roxbury Publishing Company, S. 11–22.
- Paternoster, Raymond; Bachman, Ronet (2001b): Control Theories of Crime. In: Paternoster, Raymond; Bachman, Ronet (Hg.): *Explaining Criminals and Crime*. Los Angeles, CA: Roxbury Publishing Company, S. 73–80.
- Rabe-Hesketh, Sophia; Skrondal, Anders (2012): *Multilevel and Longitudinal Modeling Using Stata*, 3. Auflage. College Station, TX: Stata Press.

- Raudenbush, Stephen W.; Bryk, Anthony S.; Cheong, Yuk F.; Congdon, Richard und du Toit, Mathilda (2004): HLM6: Hierarchical Linear and Nonlinear Modeling. Lincolnwood, IL: Scientific Software International.
- Sampson, Robert J.; Raudenbush, Stephen W. und Earls, Felton (1997): Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy. In: *Science*, 277, S. 918–924.
- Sapouna, Maria (2010): Collective efficacy in the school context: does it help explain victimization and bullying among Greek primary and secondary school students? In: *Journal of Interpersonal Violence*, 25, S. 1912–1927.
- Shaw, Clifford R.; McKay, Henry D. (1969): *Juvenile Delinquency in Urban Areas*, überarbeitete Auflage. Chicago: University of Chicago Press.
- Snijders, Tom A. B. (2005): Power and Sample Size in Multilevel Modelling. In: Everitt, B. S.; Howell, D. C. (Hg.): *Encyclopedia of Statistics in Behavioral Science*. Band 3. Chichester, West Sussex: Wiley, S. 1570–1573.
- Snijders, Tom A. B.; Bosker, Roul J. (2012): *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*, 2. Auflage. London: Sage.
- StataCorp (2013): *Stata Base Reference Manual*. Release 13. College Station, TX: StataCorp LP.
- Windzio, Michael (2008): Social Structures and Actors: The Application of Multilevel Analysis in Migration Research. In: *Romanian Journal of Population Research*, 2, S. 113–138.
- Windzio, Michael (2013): *Regressionsmodelle für Zustände und Ereignisse. Eine Einführung*. Wiesbaden: VS Verlag.
- Windzio, Michael; Baier, Dirk (2009): Violent Behavior of Juveniles in a Multiethnic Society: Effects of Personal Characteristics, Urban Areas, and Immigrants' Peer Networks. In: *Journal of Ethnicity in Criminal Justice*, 7, S. 237–270.