Gesundheitsberichterstattung

Die Berechnung von Impfintervallen für Einschulungsuntersuchungen

von **Sebastian Beil**

Die im Rahmen der Berliner Einschulungsuntersuchungen erhobenen Impfdaten gehen in die bezirkliche, berlinweite und bundesweite Gesundheitsberichterstattung ein. In den einschlägigen Berichten werden die beobachteten Impfquoten ausgewiesen, fehlende Werte werden bei der Berechnung von Kennzahlen weitestgehend vernachlässigt. Mit dem Anteil fehlender Werte steigt jedoch auch die Unsicherheit über die tatsächlichen Impfquoten in der Grundgesamtheit. In Friedrichshain-Kreuzberg variierte der Anteil der Kinder ohne Impfpass in den Jahren 2014–2017 je nach Herkunftsgruppe zwischen 5 % und 17 %. Der Artikel zeigt, wie mithilfe einfacher Methoden und unter Hinzunahme nachvollziehbarer Annahmen Impfintervalle berechnet werden können.

Einführung

Im Jahr 2017 wurden in Berlin circa 31500 Vorschulkinder erstmalig von den bezirklichen Kinder- und Jugendgesundheitsdiensten untersucht (vgl. Bettge & Oberwöhrmann 2019, Tab. 2.3). Diese Einschulungsuntersuchung (ESU) ist gesetzlich vorgeschrieben und erfolgt in erster Linie mit dem Ziel einer besseren Vorbereitung auf den Übergang in die Schule.[1] Mithilfe eines Entwicklungsscreenings sollen mögliche Entwicklungsverzögerungen oder -störungen identifiziert und unterstützende Maßnahmen begründet werden. In wenigen Fällen wird, meist auf Antrag der Erziehungsberechtigten, das Kind für ein Jahr vom Schulbesuch zurückgestellt. Neben dieser Entwicklungsdiagnostik kommen auch andere Untersuchungsmethoden zum Einsatz, so unter anderem die Erhebung von Körpergewicht und -größe, ein Seh- und Hörtest sowie eine soziale Anamnese zu den Lebensverhältnissen der Kinder. Die Erhebungsmethoden sind stark standardisiert und die Durchführung erfolgt durch erfahrene Untersucherinnen und Untersucher (vgl. Bettge et al. 2006, Bettge et al. 2019 sowie Oldenhage et al. 2009).

Die im Rahmen der ESU erhobenen Daten sind darüber hinaus Grundlage für die Gesundheitsberichterstattung der Bezirke, des Landes Berlin und des Bundes. Die gegenwärtige Praxis sieht vor, dass die erhobenen Daten von den bezirklichen Kinder- und Jugendgesundheitsdiensten an die zuständige Senatsverwaltung für Gesundheit, Pflege und Gleichstellung (SenGPG) übermittelt, dort aufbereitet und (erstmals) ausgewertet werden und dann als Mikrodaten der bezirklichen Sozial- und Gesundheitsberichterstattung sowie in aggregierter Form dem Robert Koch-Institut für die Berichterstattung über die Impfsituation überstellt werden (vgl. Robert Koch-Institut 2018).

Erhebung des Impfstatus

In der Einladung zur ESU werden die Eltern gebeten, den Impfpass zur Untersuchung mitzubringen. Der Impfstatus des Kindes wird dann anhand der Einträge im Impfpass erhoben. Bei der Erfassung der Impfungen wird mit einem einheitlichen Meldebogen gearbeitet, den die Bundesländer mit dem Robert Koch-Institut abgestimmt haben (vgl. Robert Koch-Institut 2018). Dies ist erforderlich, da die Wahrscheinlichkeit für eine erfolgreiche Immunisierung gegen die verschiedenen Erkrankungen sowohl vom Alter des Kindes bei der ersten Impfung als auch von der Art des Impfstoffs (Mehrfach- versus Einfachimpfstoff) und den zeitlichen Abständen zwischen den Impfungen abhängt.

Sofern der Impfpass nicht vorgelegt werden kann, wird bei der Berliner ESU (und auch in Baden-Württemberg) nach dem Grund des Fehlens gefragt. Kinder ohne Impfpass, deren Eltern glaubhaft darstellen können, dass das Kind keinerlei Impfungen erhalten hat, werden mit dem Impfstatus „keine Impfung erfolgt“ erfasst. Alle anderen Kinder ohne Impfpass erhalten fehlende Werte auf allen Impfvariablen und werden von der Auswertung ausgeschlossen.

1 Rechtsgrundlagen sind § 5 der Verordnung über den Bildungsgang der Grundschule (Grundschulverordnung – GsVO) vom 19. Januar 2005 (GVBl. 2005, 16, 140) und § 34 des Gesetzes zur Verhütung und Bekämpfung von Infektionskrankheiten beim Menschen (Infektionsschutzgesetz – IfSG) vom 20. Juli 2000 (BGBl. I S. 1045), das zuletzt durch Artikel 3 des Gesetzes vom 27. März 2020 (BGBl. I S. 587) geändert worden ist.

Wann fehlende Werte ignoriert werden können

Der Ausschluss von Untersuchungseinheiten mit fehlenden Angaben bei der Berechnung von Kennzahlen ist unproblematisch, wenn angenommen werden kann, dass das Fehlen der Werte (hier: der Impfpässe) zufällig zustande gekommen ist und nicht im Zusammenhang mit den tatsächlichen Merkmalsausprägungen der zu beschreibenden Variable steht. Der Ausfallmechanismus *(missing-data-mechanism)*, der zum Fehlen der Werte geführt hat, wird dann als *missing completely at random* (MCAR) bezeichnet (vgl. Enders 2010). Trifft die MCAR-Annahme zu, dann können die beobachteten Werte als einfache Zufallsauswahl aus allen (beobachteten und unbeobachteten) Werten aufgefasst werden. Der Ausschluss der fehlenden Werte wäre in diesem Fall nicht mehr als ein Problem kleiner(er) Fallzahlen. Bei Vollerhebungen wie der ESU müssten bei einem entsprechend hohen Anteil an fehlenden Werten jedoch auch für die Berechnung der Kennzahlen inferenzstatistische Methoden eingesetzt werden, auf deren Einsatz bei vollständiger Beobachtung aller Werte verzichtet werden könnte.[2]

Mögliche Ausprägungen des Impfstatus

D	Y'	Y
0	m	1
0	m	0
1	1	1
1	0	0

Fehlende Werte machen die Unterscheidung zwischen dem beobachteten Impfstatus Y' und dem tatsächlichen Impfstatus Y notwendig. Die Variable Y' (beobachteter Impfstatus) soll der Einfachheit halber die drei Ausprägungen 0 „keine Impfung erhalten", 1 „Impfung erhalten" und „m" „nicht beobachtet" (*missing*) haben ($Y' = [0,1,m]$). Der tatsächliche Impfstatus Y hat hingegen nur die beiden Ausprägungen 0 „keine Impfung erhalten" und 1 „Impfung erhalten". Ob der Impfpass zur ESU vorlag, wird über eine Indikatorvariable $D = [0,1]$ angezeigt (vgl. zu diesem Vorgehen Enders 2010). Unter der Annahme, dass die Impfung korrekt dokumentiert wurde, gilt für Kinder mit vorliegendem Impfpass ($D = 1$): $Y'_{(D=1)} = Y_{(D=1)}$. Für Kinder ohne Impfpass gilt hingegen $Y'_{(D=0)} = m \neq Y_{(D=0)} = [0,1]$. Insgesamt sind die in der kleinen Tabelle (links) veranschaulichten Kombinationen von Ausprägungen möglich.

Da der beobachtete Impfstatus für Kinder ohne Impfpass nicht ihrem tatsächlichen Impfstatus entspricht ($Y'_{(D=0)} \neq Y_{(D=0)}$), kann die Impfquote $W(Y = 1)$ (kurz: $W[Y]$) nicht allein auf Grundlage der beobachteten Daten berechnet werden. Um dies zu verdeutlichen, soll $W(Y)$ in ihre Einzelteile zerlegt werden (vgl. Manski 1995, S. 23):

$$W(Y) = W(Y|D = 1) \cdot W(D = 1) + W(Y|D = 0) \cdot W(D = 0),$$

wobei mit $W(Y|D)$ die bedingte Wahrscheinlichkeit von Y gegeben D ist. Die Wahrscheinlichkeiten für beobachtete und fehlende Werte ($W[D = 1]$ und $W[D = 0]$) können aus den Daten über die entsprechenden Anteilswerte berechnet (bzw. im Fall von Stichprobendaten „geschätzt") werden. Unter der oben getroffenen Annahme, dass die Impfung im Impfpass korrekt dokumentiert wurde, kann auch die Impfquote der Kinder, deren Impfpass zur ESU vorlag, berechnet werden ($W[Y|D = 1]$). Die Impfquote der Kinder ohne vorgelegten Impfpass ($W[Y|D = 0]$) ist hingegen nicht identifiziert.

An dieser Stelle „hilft" die MCAR-Annahme, denn nach ihr gilt (zumindest asymptotisch):

$$W(Y) = W(Y|D = 1) = W(Y|D = 0).$$

Unter der MCAR-Annahme können die beobachteten Daten demnach auf alle Untersuchungseinheiten generalisiert werden.

Was tun, wenn der Ausfallmechanismus nicht ignoriert werden kann?

Es ist davon auszugehen, dass die MCAR-Annahme für den vorliegenden Anwendungsfall nicht haltbar ist. So kommt beispielsweise das Robert Koch-Institut (2018, S. 155) nach Sichtung der Studienlage zum Ergebnis, dass „[d]ie auf der Basis der vorgelegten Impfausweise berechneten Impfquoten […] vermutlich eine leichte Überschätzung der erzielten Impfquoten dar[stellen]". Welches Ausmaß diese Überschätzung hat und ob für bestimmte soziale Gruppen auch eine Unterschätzung möglich ist, ist bislang nicht systematisch erforscht worden.

Bedingt ignorierbarer Ausfallmechanismus (*missing at random*)

Entscheidend für die Zuordnung einer Datenstruktur mit fehlenden Werten zu einem Ausfallmechanismus ist der Zusammenhang zwischen D,Y und anderen Merkmalen. MCAR setzt voraus, dass kein (bedingter oder unbedingter) Zusammenhang zwischen D und Y besteht (es gilt: $D \perp Y$, vgl. das erste Panel in Abbildung a).[3] Wird die MCAR-Annahme verworfen, muss der Ausfallmechanismus (das heißt die Entstehung der fehlenden Werte) zunächst weiter ergründet werden. Ist das Fehlen der Werte allein auf Prozesse zurückzuführen, die mithilfe von beobachteten Variablen (X) abgebildet werden können, so wäre es möglich, diese Variablen bei der Daten-

2 Aufgrund des in der Regel sehr großen Auswahlsatzes (das heißt des geringen Anteils fehlender Werte) werden die berechneten Konfidenzintervalle sehr klein sein. Ob bei Vollerhebungen inferenzstatistische Methoden eingesetzt werden müssen, hängt von den Zielparametern ab. Soll die mittels Vollerhebung erfasste Grundgesamtheit nur beschrieben werden, so ist in der Regel keine schließende Statistik notwendig. Soll hingegen auf den datengenerierenden Prozess geschlossen werden, der die Grundgesamtheit hervorgebracht hat, so kann dies sinnvoll sein.

3 Das bedeutet jedoch nicht, dass kein Zusammenhang zwischen Variablen in X und der Variable Y bestehen darf. Auch ist es möglich, dass Variablen in Z zwar D, nicht jedoch gleichzeitig Y beeinflussen. Wichtig ist lediglich, dass durch diese Zusammenhänge keine statistische Beziehung zwischen D und Y entsteht.

1 | Impfquoten für Masern und fehlende Werte 2014 bis 2017 in Friedrichshain-Kreuzberg nach kultureller Herkunft

Kulturelle Herkunft[1]	Grundimmunisierung					
	(1)	(2)	(3)	(4)	(5)	(6)
	keine Impfung	1 Impfdosis	2+ Impfdosen	n	N	Miss
	%			Anzahl		%
deutsch	*6*	*6*	*88*	4 051	4 263	*5*
türkisch	*1*	*3*	*96*	1 490	1 587	*6*
arabisch	*3*	*6*	*92*	719	818	*12*
osteuropäisch	*7*	*11*	*83*	748	899	*17*
westliche Industriestaaten	*9*	*7*	*84*	790	867	*9*
sonstige Staaten	*3*	*7*	*90*	713	814	*12*
Gesamtquote	*5,0*	*5,9*	*89,1*	8 511	9 259	*8*
bedingte Gesamtquote	*5,0*	*6,0*	*89,0*	9 259	–	–

1 57 fehlende Werte in der Variable „kulturelle Herkunft"; Spalten (1)–(3): nur Kinder mit Impfpass oder ohne Impfpass und ohne jegliche Impfung; zusätzlich zur Zahl der gültigen Werte (4) wird hier der Anteil fehlender Werte an allen Werten angegeben (6)

Quelle: ESU-Daten von SenGPG, eigene Berechnungen

analyse oder aber in einem Modell zur Imputation[4] der fehlenden Werte zu berücksichtigen. In diesem Fall gilt der Ausfallmechanismus zumindest theoretisch als bedingt ignorierbar (*missing at random*, MAR), da der Zusammenhang zwischen *D* und *Y* verschwindet, sobald in geeigneter Weise für *X* kontrolliert wird (vgl. Panel 2 in Abbildung a).

In den Daten der ESU (2014–2017, Bezirk Friedrichshain-Kreuzberg) gibt es beispielsweise einen statistischen Zusammenhang zwischen dem Anteil der fehlenden Impfangaben und einer als „kulturelle Herkunft" bezeichneten Variablen (vgl. Tabelle 1). Die kulturelle Herkunft wird anhand der Angaben zur Staatsangehörigkeit und zum Geburtsland der Eltern sowie zum Geburtsland des Kindes gebildet.

Wie in Tabelle 1 zu erkennen ist, schwankte der Anteil der fehlenden Werte ($W[D=0]$) in den Jahren 2014 bis 2017 im Bezirk zwischen 5 % für die herkunftsdeutschen Kinder und 17 % für „osteuropastämmige" Kinder.[5] Gleichzeitig variiert auch der beobachtete Anteil der einmal geimpften Kinder ($W[Y=1|D=1]$) bzw. der zweimal oder öfter geimpften Kinder ($W[Y=2|D=1]$) über die Herkunftsgruppen.[6] Der Anteil der Kinder mit Impfpass und vollständiger Grundimmunisierung gegen Masern (zwei und mehr Impfungen) betrug mindestens 83 % (osteuropastämmige Kinder) und höchstens 96 % (türkeistämmige Kinder). Der Anteil der Kinder mit Impfpass und einer Impfdosis gegen Masern schwankte zwischen 3 % (türkeistämmige Kinder) und 11 % (osteuropastämmige Kinder).

Würde angenommen, dass die fehlenden Werte innerhalb der durch die kulturelle Herkunft beschriebenen Gruppen zufällig entstanden sind (MAR-Annahme), so könnten die beobachteten herkunftsspezifischen Impfquoten ($W[Y|D=1,X=x]$, wobei *X* ein Indikator für die verschiedenen Gruppen ist) generalisiert und als tatsächliche herkunftsspezifische Impfquoten ($W[Y|X=x]$) aufgefasst werden.[7] Zu beachten ist jedoch, dass die in der vorletzten Zeile aufgeführte Gesamtquote ($W[Y|.]$, Randverteilung der Kreuztabelle) die Variation im Anteil der fehlenden Werte und in den Impfquoten ignoriert. Bei der Berechnung einer bedingten Gesamtquote aus den herkunftsspezifischen Einzelquoten müsste unter anderem dem Umstand Rechnung getragen werden, dass 17 % der osteuropastämmigen Kinder mit fehlenden Impfangaben eine („geschätzte") Grundimmunisierungsquote (zwei und mehr Impfungen) von 83 % haben, während nur 6 % der türkeistämmigen Kinder mit fehlendem Impfpass mit einer Quote von 96 % in die Berechnung der Gesamtquote eingehen sollten. Die letzte Zeile enthält die derart berechnete bedingte Gesamtquote, die sich leicht von der Gesamtquote unterscheidet.

Nicht ignorierbarer Ausfallmechanismus (missing not at random)

In einigen Anwendungen ist selbst die etwas leichtere MAR-Annahme nicht haltbar. Der Ausfallmechanismus gilt als nicht ignorierbar *(missing not at random,* MNAR), wenn das Fehlen eines Wertes ($D=0$) von der tatsächlichen Merkmalsausprägung (*Y*) oder aber unbeobachteten Merkmalen (*Z*) abhängt, die gleichzeitig auch *Y* beeinflussen (Panel 4 und 3 in Abbildung a).

a | Missing-Data-Mechanismen
(in Anlehnung an Enders 2010, S. 12)

Anmerkung: Pfeile mit zwei Enden (↔) stehen für einen durch beobachtete (*X*) oder unbeobachtete (*Z*) Drittvariablen bedingten Zusammenhang.

4 Die multiple Imputation trägt der Unsicherheit durch Ersetzen der Werte Rechnung. Neben der Imputation ist auch eine Schätzung mittels *full information maximum likelihood* (FIML) möglich (vgl. Enders 2010). Mehr zum Begriff „Imputation" im „Statistik erklärt", S. 16.

5 Der Anteil der fehlenden Werte (Impfpässe) korreliert dabei sehr stark mit dem Anteil der Kinder, die nicht in Deutschland geboren wurden. Dieser Anteil ist unter herkunftsdeutschen und türkeistämmigen Kindern sehr niedrig und unter osteuropastämmigen Kindern sehr hoch.

6 Wird wie hier zwischen einer einmaligen Impfung und der vollständigen Grundimmunisierung (zwei und mehr Impfungen) unterschieden, so hat *Y* drei Ausprägungen ($Y=[1,2,3]$).

7 Dieses Verfahren ähnelt der *conditional mean* oder *regression imputation* (vgl. Enders 2010, S. 44 ff.).

2 | No-assumption-bounds und gedeckelte bounds für Masernimpfung 2014 bis 2017 in Friedrichshain-Kreuzberg nach kultureller Herkunft

Kulturelle Herkunft[1]	Impfung gegen Masern		n	Miss	no assumption bound für $Y=1$		gedeckelter bound für $Y=1$		N
	$Y=0$	$Y=1$							
	keine	mind. 1			Unter-	Ober-	Unter-	Ober-	
	Impfung				grenze				
	%		Anzahl	%					Anzahl
deutsch	*6,1*	*93,9*	4 051	5	*89,2*	*94,2*	*89,2*	*93,9*	4 263
türkisch	*0,9*	*99,1*	1 490	6	*93,1*	*99,2*	*93,1*	*99,1*	1 587
arabisch	*2,9*	*97,1*	719	12	*85,3*	*97,4*	*85,3*	*97,1*	818
osteuropäisch	*6,6*	*93,5*	748	17	*77,8*	*94,5*	*77,8*	*93,5*	899
westliche Industriestaaten	*8,9*	*91,1*	790	9	*83,0*	*91,9*	*83,0*	*91,1*	867
sonstige Staaten	*3,0*	*97,1*	713	12	*85,0*	*97,4*	*85,0*	*97,1*	814
Gesamt	*5,0*	*95,0*	8 511	8	*87,5*	*95,4*	*87,5*	*95,0*	9 259

1 57 fehlende Werte in der Variable „kulturelle Herkunft".

Quelle: ESU-Daten von SenGPG, eigene Berechnungen.

Das Robert Koch-Institut (2018) geht davon aus, dass Kinder ohne Impfpass weniger oft geimpft sind als Kinder mit Impfpass. Für Kinder, die in anderen Ländern geboren wurden, liegt häufiger kein (auswertbarer) Impfpass zur ESU vor. Gründe hierfür können der Verlust des Impfpasses bei der (unfreiwilligen) Ausreise, eine fehlende Übersetzung der Impfangaben, aber auch die unterschiedliche Gesundheitsversorgung in den Herkunftsländern und der erschwerte Zugang zur Gesundheitsversorgung im Gastland sein (vgl. Robert Koch-Institut, 2012).

Während Migrationserfahrungen der Eltern beziehungsweise Kinder teilweise durch Variablen im Datensatz abgebildet (und damit prinzipiell kontrolliert) werden können, gilt dies für andere Merkmale nicht. So könnten Eltern, die mit ihren Kindern selten oder nicht an Vorsorgemaßnahmen teilnehmen, gegebenenfalls aus Scham versuchen, der sozialen (ärztlichen) Kontrolle zu entgehen und deshalb keinen Impfpass zur ESU mitbringen. Ganz allgemein könnte vermutet werden, dass Eltern, die vergessen, den Impfpass zur Untersuchung mitzubringen, bzw. diesen nicht auffinden können, Vorsorgeuntersuchungen und Impfungen weniger Aufmerksamkeit beimessen als Eltern, die den Impfpass mitbringen.[8] Da die Gründe für das Vorlegen des Impfpasses nicht (ausreichend) erfasst werden, kann über den Impfstatus der Kinder mit fehlendem Impfpass letztlich nur (mehr oder weniger gut) spekuliert werden. Erhoben wird nur, ob keine Impfungen erfolgt sind.

Der Unsicherheit Rechnung tragen, oder: vom Punkt zum Intervall

Wenn, wie im vorliegenden Fall, der Verdacht besteht, dass der Ausfallmechanismus nicht (bedingt) ignorierbar ist, so muss auf andere Analyseverfahren zurückgegriffen werden. Während Selektionsmodelle weiterer (oft sehr strikter) Annahmen bedürfen, kann mithilfe von Intervallschätzungen auch gänzlich auf Annahmen verzichtet werden. Diese Intervalle *(bounds)* tragen der Unsicherheit des datengenerierenden Prozesses Rechnung, sind jedoch von Konfidenzintervallen zu unterscheiden. Während Konfidenzintervalle die Unsicherheit bei der Stichprobenziehung abbilden sollen, können mithilfe von *bounds* Unsicherheiten durch fehlende Werte abgebildet werden (vgl. Manski 1995). Diese fehlenden Werte entstehen beispielsweise durch Item- oder Unit-Nonresponse oder durch das „grundlegende Problem der Kausalanalyse" (vgl. Holland 1986).[9]

Ausgangspunkt für die weitere Diskussion soll eine Variante von Tabelle 1 sein, in der einfachheitshalber nicht zwischen einer und mehreren Impfungen unterschieden wird (vgl. Tabelle 2). Die Untergrenze des einfachsten Intervalls (dem sogenannten *no-assumption-bound*) wird unter der Annahme berechnet, dass keines der Kinder mit fehlenden Impfangaben geimpft wurde ($W[Y|D=0,X=x]=0$). Die Obergrenze dieses Intervalls berechnet sich demnach aus der gegensätzlichen Annahme einer vollständigen Impfung aller Kinder ohne Impfangaben ($W[Y|D=0,X=x]=1$):

$$W(Y=1|D=1,X=x)\cdot W(D=1|X=x) \leq W(Y=1|X=x) \leq W(Y=1|D=1,X=x)\cdot W(D=1|X=x)+W(D=0|X=x)$$

Dieser no-assumption-bound ist stets so breit wie der Anteil der fehlenden Werte in der jeweiligen Herkunftsgruppe ($W[D=0|X=x]$).

Tabelle 2 kann entnommen werden, dass der Anteil der türkeistämmigen Kinder, die in den Jahren 2014 bis 2017 im Bezirk Friedrichshain-Kreuzberg vom Kinder- und Jugendgesundheitsdienst (KJGD) untersucht wurden und die mindestens eine Masernimpfung erhalten haben, im Intervall von 93,1 %

8 Einige Untersuchende lassen sich den Impfpass nachreichen.

9 Das grundlegende Problem der Kausalanalyse besteht darin, dass eine Untersuchungseinheit zu einem Zeitpunkt nur in einem Zustand beobachtet werden kann, weshalb der Effekt eines Zustands auf ein bestimmtes Merkmal entweder durch den Vergleich verschiedener Individuen in unterschiedlichen Zuständen (Querschnittsdaten) oder durch Beobachtung von Individuen über die Zeit (Paneldaten) geschätzt werden muss. Paneldaten ermöglichen den intraindividuellen Vergleich, jedoch muss für andere zeitveränderliche Variablen kontrolliert werden.

Statistik erklärt: Statistische Imputation

In statistischen Untersuchungen wird meist ein „rechteckiger Datensatz" beziehungsweise eine Datenmatrix analysiert. Dabei beschreiben die Zeilen dieser Matrix die Einheiten respektive Beobachtungen; im Englischen auch als „*units*" respektive „*cases*" bezeichnet. Die Spalten einer Datenmatrix geben die Variablen einer statistischen Erhebung wieder, die für jede der Einheiten erhoben werden. Aus unterschiedlichen Gründen kann es vorkommen, dass die Datenmatrix nicht vollständig ist und einzelne Daten fehlen – ganze Beobachtungen (Zeilen) oder auch einzelne Merkmale in den Beobachtungen. Die Gründe dafür können ganz unterschiedlich sein. Für die spätere Verwendung der Daten gibt es verschiedene Methoden, mit solchen Antwortausfällen umzugehen. Diese Methoden werden unter dem Begriff „*Missing-Data*-Techniken" zusammengefasst. Beispielsweise gehören Eliminierungsverfahren („*Complete-case analysis*") zu diesen Techniken. Dabei werden alle unvollständigen Einheiten, die einen oder mehrere fehlende Werte aufweisen, aus der Analyse ausgeschlossen. Dies ist zwar eine technisch sehr einfache Methode, bringt jedoch einige Nachteile mit sich: Zum einen kann der Stichprobenumfang auf diese Weise extrem klein werden, womit viele wertvolle Informationen verloren gehen. Zum anderen besteht die Möglichkeit, dass die Analyseergebnisse nach der Anwendung von Eliminierungsverfahren Verzerrungen enthalten. Wenn etwa in einer Umfrage Viel-Verdiener statistisch seltener eine Gehaltsangabe machen und diese Fälle anschließend ignoriert werden, ist die Repräsentativität der Umfrage nicht mehr gegeben. Dieses Problem tritt insbesondere dann auf, wenn der Ausfallmechanismus nicht rein zufällig und damit ignorierbar („*missing not at random*") ist. In solchen Fällen ist es ratsam fehlende Daten zu imputieren, also die Antwortausfälle durch plausible Werte zu ersetzen.

Im Allgemeinen werden bei der Imputation die fehlenden Werte mittels eines entsprechenden Imputationsverfahrens unter Berücksichtigung der beobachteten Werte des gleichen Datensatzes geschätzt und dann die Kennwerte auf Basis des vervollständigten Datensatzes berechnet. Am obigen Beispiel könnte den Personen ohne Gehaltsangabe der Mittelwert ähnlicher Fälle (zum Beispiel gleiches Alter, Beruf, Wohnviertel) zugeordnet werden.

Der Begriff „Imputation" bezeichnet dabei jedoch nicht ein einzelnes Verfahren, sondern eine Vielzahl unterschiedlicher Methoden. Die lassen sich grob in die singuläre (oder Einfach-)Imputation und die multiple Imputation einteilen. Während die fehlenden Werte bei der Einfachimputation nur einmal imputiert werden, werden bei der multiplen Imputation mehrere Datensätze mit verschiedenen imputierten Werten generiert und die Datensätze im Anschluss gemeinsam ausgewertet.

bis 99,2 % gelegen hat. Dieses Intervall überschneidet sich nicht mit dem Intervall für Kinder aus den westlichen Industriestaaten, weshalb davon ausgegangen werden kann, dass türkeistämmige Kinder in Friedrichshain-Kreuzberg besser geimpft sind als Kinder aus westlichen Industriestaaten. Die berechneten Intervalle beruhen einzig auf den relativ schwachen Annahmen, dass 1. die Impfungen korrekt in den vorliegenden Impfpässen dokumentiert wurden und 2. die Daten aus den vorliegenden Impfpässen korrekt (das heißt, ohne systematische Fehler) in das Erhebungsprogramm übertragen wurden. Annahmen über das Entstehen der fehlenden Werte müssen hingegen nicht getroffen werden.

Wie bereits beschrieben, gehen Forscherinnen und Forscher des Robert Koch-Instituts (2018) davon aus, dass Kinder ohne Impfpass weniger oft geimpft sind als Kinder mit Impfpass. Wird diese Annahme für alle Herkunftsgruppen als gleichermaßen plausibel betrachtet, so lassen sich die Impfintervalle verkleinern *(gedeckelte bounds)*.

$$W(Y=1|D=1,X=x)\cdot W(D=1|X=x) \leq W(Y=1|X=x) \leq W(Y=1|D=1,X=x)\cdot[W(D=1|X=x)+W(D=0|X=x)]$$

Die Obergrenze entspricht nun der beobachteten Impfquote, während für die Untergrenze weiterhin gilt, dass kein Kind ohne Impfpass geimpft wurde. Während sich das Impfintervall für die türkeistämmigen Kinder aufgrund der ohnehin sehr hohen Impfquote und des geringen Anteils an fehlenden Werten nur sehr leicht verändert, schrumpft es für die osteuropastämmigen Kinder und die Kinder aus den westlichen Industriestaaten um einen Prozentpunkt. Über alle untersuchten Kinder ist das berechnete Impfintervall 7,5 Prozentpunkte breit.

Jede weitere Annahme, die die Variation des unbekannten Anteils geimpfter Kinder unter den Kindern ohne Impfpass begrenzt, würde auch zu einer Verkleinerung der Impfintervalle führen. Diese Annahmen können sich auf theoretische Überlegungen oder aber empirische Beobachtungen aus anderen Studien (verknüpft mit Annahmen über die Übertragbarkeit der Ergebnisse auf die eigene Untersuchung) stützen. Diese vermeintlich „bessere" Schätzung (also die Verkleinerung der Impfintervalle) hat jedoch einen Preis: Die durch die Datenerhebung entstandene Unsicherheit (fehlende Werte) wird gegen die Unsicherheit der getroffenen Annahmen eingetauscht.

Fazit

Die meisten Erhebungen sind von Item- oder Unit-Nonresponse betroffen. Oft werden die dadurch entstandenen fehlenden Werte ignoriert, ohne die für diese Analysestrategie notwendige Annahme *(missing completely at random)* zu überprüfen. Für die fehlenden Impfangaben in den Einschulungsuntersuchungsdaten trifft diese Annahme nicht zu, weshalb für einen anderen Umgang mit den fehlenden Werten plädiert wurde. Sofern, wie im vorliegenden Fall, nicht ausgeschlossen werden kann, dass unbeobachtete Variablen (wie bspw. eine impfskeptische Haltung, Zugangsbarrieren oder auch elterliche Fürsorge bzw. Disziplin) sowohl den Impfstatus als auch das Fehlen des Impfpasses beeinflussen, können Impfintervalle *(bounds)* berechnet werden, die der dadurch entstehenden Unsicherheit Rechnung tragen. Die Berechnung von Impfintervallen ist relativ einfach und erfordert keine starken Annahmen. Auch die Einbindung zusätzlicher Annahmen, die eine stärkere identifizierende Wirkung entfalten (und die Intervalle verkleinern), ist einfach und vor allem transparent.

Dr. Sebastian Beil ist im Bezirksamt Friedrichshain-Kreuzberg von Berlin zuständig für die sozialwissenschaftliche Datenerhebung und -analyse sowie Berichterstattung. Als Teil der Organisationseinheit für Bezirkliche Planung und Koordinierung ist er darüber hinaus mit der Organisation und Durchführung von Beteiligungsverfahren sowie der Abstimmung von Fachplanungen befasst.
Am 23. Oktober 2019 stellte er im Rahmen eines statistischen Kolloquiums am Standort Berlin des Amtes für Statistik Berlin-Brandenburg Ergebnisse der Einschulungsuntersuchungen zum Impfstatus der Kinder in Friedrichshain-Kreuzberg vor.

Literaturverzeichnis

- Bettge, S.; Oberwöhrmann, S. (2019): Grundauswertung der Einschulungsdaten in Berlin 2017. Berlin: Senatsverwaltung für Gesundheit, Pflege und Gleichstellung.
- Bettge, S.; Oberwöhrmann, S.; Delekat, D.; Häßler, K.; Herrmann, S.; Meinlschmidt, G. (2006): Zur gesundheitlichen und sozialen Lage von Kindern in Berlin. Spezialbericht, Berlin.
- Enders, C. (2010): Applied Missing Data Analysis. New York: The Guilford Press.
- Holland, P. W. (1986): Statistics and Causal Inference. Journal of the American Statistical Association, S. 945–960.
- Manski, C. (1995): Identification Problems in the Social Sciences. Cambridge: Harvard University Press.
- Oldenhage, M.; Daseking, M.; Petermann, F. (2009): Erhebung des Entwicklungsstandes im Rahmen der ärztlichen Schuleingangsuntersuchung. Gesundheitswesen, S. 638–647.
- Robert Koch-Institut (2012). Zugangsbarrieren, Zugangswege und Ressourcen, https://www.rki.de/DE/Content/Infekt/Impfen/Migration/Zugangswege/migration_zugangswege_node.html, abgerufen am 20.01.2020.
- Robert Koch-Institut (2018): Impfquoten bei der Schuleingangsuntersuchung in Deutschland 2016. Epidemiologisches Bulletin, Nr. 16/2018, S. 151–156.