

Verglichen womit? Kontrollen für Fall-Kontroll-Studien finden

David A. Grimes und Kenneth F. Schulz

Family Health International, PO Box 13950, Research, Triangle Park, NC 27709, USA

aus: *Lancet* 2005; **365**: 1429–33: (Originaltitel: „Compared to what? Finding controls for case-control studies“)

Zusammenfassung

Die Verwendung von Kontroll- oder Vergleichsgruppen ist ein leistungsfähiges wissenschaftliches Instrument. In Fall-Kontroll-Studien schätzt man über Kontrollen die Expositionshäufigkeit in der zugrunde liegenden Population. Kontrollen können bekannten oder unbekanntem Studienpopulationen entstammen. Eine bekannte Gruppe besteht aus einer definierten Grundgesamtheit, die über einen bestimmten Zeitraum beobachtet wird, beispielsweise den Passagieren eines Kreuzfahrtschiffes. Ist die Studiengruppe bekannt, so kann als Kontrolle eine aus dieser Population gezogene Stichprobe dienen. Liegt für die Grundgesamtheit keine Namensliste vor, bieten sich Verfahren wie die telefonische Zufallsziffernanwahl (Random Digit

Dialling) an. Wenn die Studiengruppe nicht bekannt ist (z. B. bei Verkehrsunfallopfern in der Notaufnahme, die unter Umständen aus weiter entfernten Orten kommen), bieten sich Kontrollen aus dem Krankenhaus oder aus der Nachbarschaft oder Freunde, Begleitpersonen bzw. Angehörige an. Im Allgemeinen ist eine gut selektierte Kontrollgruppe besser als zwei oder mehr Kontrollgruppen. Bei kleinen Stichproben kann man das Verhältnis von Fällen zu Kontrollen erhöhen, um die Chance zur Feststellung relevanter Unterschiede zu verbessern. Auch wenn es die ideale Kontrollgruppe nicht gibt, muss der Leser sehr genau überlegen, wie repräsentativ die Kontrollen sind. Eine schlecht ausgewählte Kontrollgruppe kann nicht nur zu falschen Ergebnissen führen, sondern aus medizinischer Sicht sogar Schaden anrichten.

se „Studie“ nicht nur wegen ihrer unzureichenden Expositionsdauer, sondern auch wegen der wahrscheinlich göttlichen Begleitintervention in die Kritik geraten. Jedenfalls sind Kontrollgruppen erst einmal für mehrere Jahrtausende aus der Literatur verschwunden!

An James Linds Untersuchung über die Behandlung von Skorbut aus dem Jahre 1747 hat sich das Interesse an gleichzeitig beobachteten Kontrollgruppen erneut entzündet [2]. Trotz ihrer geringen Größe (sechs Behandlungsgruppen mit je zwei Seeleuten) zeigte die Studie, wie nützlich die Nahrungsergänzung mit Zitrusfrüchten ist. In Studien ohne Randomisierung kann das Zusammenstellen einer angemessenen Kontrollgruppe manchmal eine wahre Herausforderung darstellen. Wir werden hier die Rolle von Kontrollgruppen in Fall-Kontroll-Studien erläutern, die besonderen, mit ihrer Auswahl verbundenen Probleme beschreiben und verschiedene Implikationen erörtern.

Sinn und Zweck von Kontrollen

Kontrollen in einer Fall-Kontroll-Studie, die ein Ergebnis rückblickend bis zur Exposition verfolgt [3], geben Hinweise auf die Hintergrundhäufigkeit einer Exposition bei Personen, die nicht an der fraglichen Krankheit leiden. Die Kontrollpersonen müssen aber nicht unbe-

In einem alten Film wurde der Komiker Groucho Marx gefragt: „Groucho, how is your wife?“ Groucho witzelte: „Compared to what?“¹ Auch wenn seine Antwort („Verglichen womit?“) nach heutigem Ermessen sexistisch anmuten mag, so passt sie doch genau zu der Frage, die sich bei den Ergebnissen von Fall-Kontroll-Studien stellt, nämlich im Vergleich wozu? Denn die Validität von Schlussfolgerungen hängt davon ab, ob sich eine geeignete Vergleichsgruppe finden lässt. Mit anderen Worten: Die Verwendung nicht so geeigneter Kontrollgrup-

pen hat schon vielen wissenschaftlichen Arbeiten die Basis entzogen. Kontrollgruppen sind ein aussagekräftiges und seit langem bewährtes Instrument der Forschung. Die allererste Dokumentation einer Vergleichsgruppe findet sich schon in der Bibel im Buch Daniel [1]. Daniel und seine drei Freunde, die vom babylonischen König Nebukadnezar gefangen gehalten wurden, nahmen am Hof des Königs an einer zehntägigen „Studie“ teil, in der gesunde Ernährung mit dem königlichen Speiseplan verglichen wurde. Am Ende schienen Daniel und seine Freunde gesünder zu sein als die babylonischen Jugendlichen, die das übliche Essen zu sich nahmen. Im Laufe der Jahre ist die-

¹ Wortspiel mit der Doppelbedeutung von „How is your wife?“: Gemeint war „Wie geht es deiner Frau?“, wörtlich verstanden wird daraus „Wie ist deine Frau?“

dingt gesund sein; gelegentlich ist auch der Einschluss Kranker angemessen. Kranke als Kontrollpersonen auszuschießen, kann die Ergebnisse verzerren [4] und somit zu einem so genannten Bias führen [5]. (Wie Gesunde können auch kranke Menschen eine andere interessierende Erkrankung entwickeln.) Der letzte Aspekt ist jedoch der wichtigste: Kontrollen in einer Fall-Kontroll-Studie sollten diejenigen repräsentieren, die unter Risiko stehen, ein „Fall“ zu werden [6]. Anders gesagt, Kontrollen sollten dasselbe Expositionsrisiko aufweisen wie die Fälle, wenn Exposition und Krankheit nicht miteinander in Beziehung stehen (s. Kasten) [7].

Wenn Fälle (Personen mit der Erkrankung) eine höhere Expositionshäufig-

Eigenschaften der Kontrollen einer Fall-Kontroll-Studie.

- frei von der interessierenden Zielerkrankung
- repräsentativ für die Personen, die unter Risiko für die Zielerkrankung stehen
- unabhängig von der interessierenden Exposition ausgewählt.

keit aufweisen als die Kontrollen (Personen ohne die Erkrankung), dann ergibt sich eine positive Assoziation. So sind etwa multiple Sexualpartner bei Fallpatientinnen mit Zervixkarzinom häufiger anzutreffen als bei den Kontrollpatientinnen ohne Zervixkarzinom. Ist die Expositionsprävalenz unter den Fällen niedriger als unter den Kontrollen, liegt eine protektive Assoziation vor – z. B. ist die Anwendung oraler Kontrazeptiva unter den Ovarialkarzinomfällen seltener zu beobachten als unter den Kontrollen ohne ein solches Karzinom.

Wichtig ist bei der Auswahl der Kontrollen für eine Fall-Kontroll-Studie die Vermeidung von systematischen Verzerrungen (Bias). Selektionsfehler entstehen, wenn die Kontrollen für Personen unter Risiko für die fragliche Erkrankung nicht repräsentativ sind. Ein gutes Beispiel sind Fall-Kontroll-Studien über potenzielle protektive Maßnahmen gegen die Entwicklung eines kolorektalen Karzinoms in Verbindung mit nichtste-

roidalen Antiphlogistika (NSAIDs) [8, 9] (siehe Abb. 1). Angenommen, die Fälle mit kolorektalem Karzinom würden zum Zeitpunkt ihrer Operation im Krankenhaus identifiziert. Dann müssten die Kontrollen Krankenhauspatienten ohne ein kolorektales Karzinom sein. Würden wir die Kontrollen aus dem rheumatologischen Bereich auswählen, würde dies unsere Ergebnisse verzerren, da Patienten mit Arthritis gegenüber NSAIDs wahrscheinlich eher exponiert sind als die meisten anderen Personen in der Bevölkerung, was den Schätzer der Assoziation zwischen NSAIDs und kolorektalem Karzinom verringern würde. Würden die Kontrollen dagegen aus dem gastroenterologischen Bereich ausgewählt, ergäbe sich dadurch eine Verzerrung der Ergebnisse in umgekehrter Richtung. Denn die Wahrscheinlichkeit einer NSAID-Exposition wäre bei Ulkuspatienten aufgrund der Warnhinweise ihrer Ärzte geringer als in der Allgemeinbevölkerung – eine Verzerrung, die zur Erhöhung des Effektschätzers führen würde [3].

Untersuchungen zur Endometriose liefern ein weiteres Beispiel dafür, welche Herausforderung die richtige Auswahl einer Kontrollgruppe darstellt. Da die Diagnose einer Endometriose operativ gestellt wird, verwenden Untersucher als Kontrollpatientinnen häufig Frauen, die sich einer Laparoskopie oder Laparotomie unterziehen, ohne dass eine Endometriose-Diagnose gestellt wird. Patientinnen mit diesen Operationen sind für die Frauen, bei denen ein Risiko für die Entstehung einer Endometriose besteht, wahrscheinlich aber nicht repräsentativ, da Operationen ja nicht zufällig durchgeführt werden [10].

Wo findet man Kontrollen?

Der Untersucher oder Prüfartz (und letztlich auch der Leser) muss die Personengruppe definieren, der die Fälle und Kontrollen entnommen werden sollen. Eine bekannte Gruppe [11] besteht aus einer definierten Population, die über einen gegebenen Zeitraum beobachtet

Auswahl der Kontrollen	Methodische Schwächen der Kontrollgruppe	Auswirkungen auf den Schätzer für den NSAID-Effekt
Patienten mit Arthritis	wahrscheinlich atypisch hohe Exposition gegenüber NSAIDs	↓
Patienten mit Ulkus	wahrscheinlich atypisch niedrige Exposition gegenüber NSAIDs	↑

Bei den Fällen handelt es sich um stationär aufgenommene Patienten mit kolorektalem Karzinom.

Abb. 1. Auftreten von Bias in einer Fall-Kontroll-Studie über nichtsteroidale Antiphlogistika (NSAIDs) und kolorektales Karzinom.

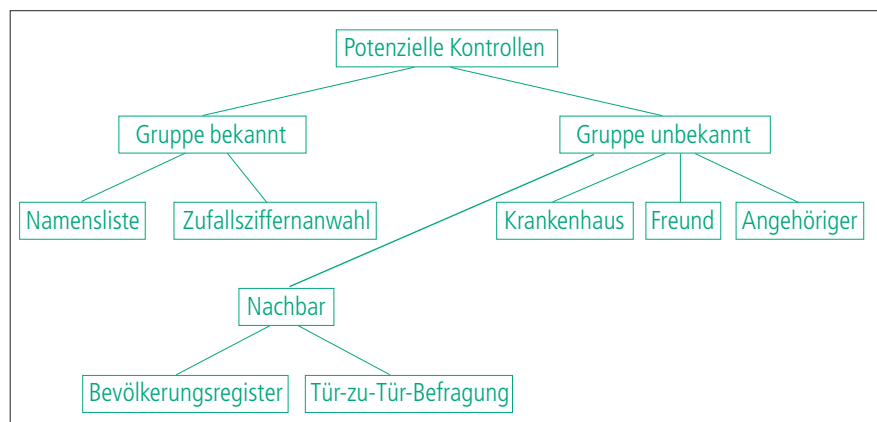


Abb. 2. Auswahl der Kontrollen in bekannten und unbekanntem Teilnehmergruppen.

wird (Abb. 2). Diese Gruppe könnte aus den Passagieren und der Schiffsbesatzung einer wochenlangen Kreuzfahrt in die Karibik bestehen oder aus allen in einem Zeitraum von zehn Jahren in Schweden lebenden Personen. Fälle sind diejenigen, die die interessierende Erkrankung entwickeln. Bei den Kontrollen handelt es sich um Personen aus derselben Gruppe ohne diese Erkrankung. Fall-Kontroll-Studien können auch in eine größere Kohortenstudie eingebettet sein. (Die sog. „Nested“-Fall-Kontroll-Studien [12] sind dafür ein schönes Beispiel.) Hier besteht die Aufgabe also darin, die Fälle in der betreffenden Gruppe zu finden; die Auswahl der Kontrollen ist in einer definierten Population leichter.

Meist ist die Gruppe, der die Fälle entstammen, unbekannt [11]. Ein Beispiel dafür sind die Opfer von Verkehrsunfällen in Krankenhausnotaufnahmen. Einige wohnen in der Nähe, andere waren vielleicht auf der Durchreise, und wieder andere wurden möglicherweise mit dem Hubschrauber aus dem ländlichen Umland eingeflogen. In dieser Situation werden die Fälle ausgewählt, bevor die Studiengruppe feststeht. Die Fälle zu finden, ist der einfache Teil der Übung; die Herausforderung besteht nun darin, die Gruppe zu definieren, aus der sich die Kontrollen rekrutieren sollen. Sie sollten derselben Gruppe entstammen. (Eine Möglichkeit wäre, Fälle und Kontrollen auf Personen zu beschränken, die innerhalb der Stadtgrenzen wohnen.)

Schlecht selektierte Kontrollgruppen können zu schwerwiegenden Fehlern führen. In einer frühen Fall-Kontroll-Studie über AIDS bei homosexuellen Männern in San Francisco wurden zwei Kontrollgruppen verwendet: eine Gruppe aus der Nachbarschaft und eine aus der Klinik [13]. Die Odds-Ratios für die Exposition (> 100 Lebenspartner) betragen 52,0 für die Nachbarschaftskontrollen und 2,9 für die Krankenhauskontrollen. Die Verwendung der Krankenhauskontrollen führte zu einer groben Unterschätzung des wahren Risikos, da die Wahrscheinlichkeit der Inanspruchnahme der Klinik in enger Beziehung zur interessierenden Exposition stand, d. h. von der Anzahl der Sexual-

partner nicht unabhängig war. Die Wahrscheinlichkeit, dass die Kontrollen aus einer Klinik für sexuell übertragbare Erkrankungen multiple Sexualpartner hatten, war sehr viel höher als bei anderen homosexuellen Männern in San Francisco.

Kontrollen aus einer bekannten Gruppe

Wenn möglich, dienen Zufallsstichproben von Personen ohne die Erkrankung als Kontrollen. Bei der Untersuchung von lebensmittelbedingten Krankheitsausbrüchen auf einem Kreuzfahrtschiff wird in der Regel ein Fall-Kontroll-Design gewählt. Als Fälle gelten die Personen, bei denen eine Gastroenteritis aufgetreten ist; die Kontrollen sind diejenigen an Bord, bei denen die Krankheit nicht ausbrach. Eine solche Studie versucht Lebensmittel zu identifizieren, die von den Fällen häufiger verzehrt wurden als von den Kontrollen. Ferner sollte niemand erkrankt sein, der das verdächtige Nahrungsmittel nicht zu sich genommen hat. Auf dem Schiff könnte aus den nicht erkrankten Personen eine Stichprobe gezogen werden [14]. Die Kontrollen könnten also eine Zufallsstichprobe aller an Bord befindlichen Personen ohne Gastroenteritis sein. Bevölkerungskontrollen haben sowohl Vor- als auch Nachteile. Die Erhebung von Zufallsstichproben sollte repräsentative Kontrollen ergeben, und die Extrapolation der Ergebnisse auf die Studiengruppe lässt sich leicht rechtfertigen. Auf der anderen Seite können

Bevölkerungskontrollen sich als ungeeignet erweisen, wenn nicht alle Fälle in der fraglichen Population identifiziert werden konnten oder wenn eine beträchtliche Anzahl der potenziellen Kontrollen nicht erreichbar ist, weil sie z. B. Ferien machen. Außerdem könnten Bevölkerungskontrollen weniger für die Teilnahme an der Untersuchung motiviert sein als Personen in medizinischen Einrichtungen, wie etwa Krankenhauspatienten [15].

Wenn für das Untersuchungskollektiv keine Namensliste vorliegt, lassen sich potenzielle Kontrollen durch Zufallsziffernwahl (Random Digit Dialling) ermitteln [7, 16]. Eine Zufallsstichprobe aus unvollständigen Telefonnummern (z. B. acht Ziffern) wird aus allen eingetragenen Telefonanschlüssen erhoben; zufällig ausgewählte zweistellige Zahlen ergeben dann die Nummer, die angerufen werden soll (Abb. 3). Ein solches Vorgehen hat sowohl Vor- als auch Nachteile.

Es wird versucht, eine gleichmäßige Stichprobe aus den privaten Telefonnummern zu ziehen und Verbindungen zu Geschäftsanschlüssen auf ein Minimum zu begrenzen. Mit dieser Strategie werden sowohl die nicht über Telefonbücher zugänglichen neuen als auch die nicht eingetragenen Telefonnummern erfasst. So wird zwar eine Zufallsstichprobe von Telefonnummern erhoben, das eigentliche Ziel ist jedoch die Erhebung einer Zufallsstichprobe potenzieller Kontrollen. Aber nicht alle Menschen haben einen Telefonanschluss; diejenigen ohne Telefon weisen tendenziell einen niedrigeren sozioökonomischen Status auf. Andere haben

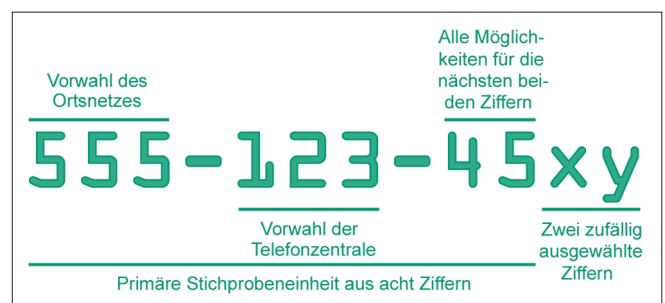


Abb. 3. Zufallsziffernwahl zur Auswahl von Kontrollen [17].

Die primäre Stichprobeneinheit umfasst achtstellige Nummern: die Vorwahlnummern aller bekannten Ortsnetzbereiche und dreistelligen Vorwahlnummern der Telefonzentralen in einem Landkreis sowie alle Kombinationsmöglichkeiten der nächsten beiden Ziffern. Für alle diese zufällig ausgewählten achtstelligen Zahlen wurden die beiden Endziffern von einem Computer generiert. Daraus entstanden zehnstellige Telefonnummern, die angerufen werden sollten.

mehr als eine Telefonnummer (z. B. Festnetzanschluss plus Mobiltelefon), was mit einem höheren sozioökonomischen Status verbunden sein könnte. Der Telefonkontakt zu diesem Personenkreis wird damit also wahrscheinlicher. Manche Telefonnummern werden von mehr als einer potenziellen Kontrolle benutzt. Außerdem unterscheiden sich Personen, die telefonische Anfragen nur ungern beantworten, von solchen, die bereitwillig an Umfragen teilnehmen [18]. So werden etwa junge Frauen seltener durch Zufallsziffernanwahl ermittelt als andere [19]. Aber auch wenn diese Eigenarten der telefonischen Erfassung zu Bias führen können [15], sind die auf diese Weise selektierten Kontrollgruppen für die Referenzpopulation doch weitgehend repräsentativ [17].

Kontrollen aus einer unbekanntem Gruppe

Nachbarschaftskontrollen

Nachbarschaftskontrollen werden im Allgemeinen nach einem festgelegten Muster aus dem Häuserblock gezogen, in dem der Fallpatient wohnt. Wie immer sollte die Selektion der Kontrollen auch hier unabhängig von der interes-

sierenden Exposition erfolgen. Zur Vermeidung von Selektionsbias gehen die Interviewer nach einem spezifischen Plan vor. Zur Identifizierung der Häuser potenzieller Kontrollen eignen sich z. B. Einwohnerregister oder Tür-zu-Tür-Befragungen [20, 21]. Für die Arbeit mit dem Einwohnerregister sind Verzeichnisse mit Querverweisen zu Anschriften und zugehörigen Telefonnummern von Nutzen.

Für eine Fall-Kontroll-Studie über orale Kontrazeptiva und hepatozelluläre Adenome entschied man sich für die Tür-zu-Tür-Befragung. Die Untersucher befragten jede Fallpatientin zu Hause und versuchten dann, in derselben Straße drei Kontrollen zu finden (Abb. 4) [22]. In einer weiteren Fall-Kontroll-Studie in einer ländlichen Gemeinde wurde die Nachbarschaft der Fallpatienten kartiert. Die Interviewer begannen mit dem Haus gegenüber und arbeiteten sich in immer weiteren Kreisen vor, bis eine passende Kontrolle gefunden war [21]. Zu den Vorteilen der Nachbarschaftskontrollen gehört, dass keine Namenslisten erforderlich und zahlreiche Störgrößen (Confounder) berechenbar sind (z. B. sozioökonomischer Status, Klima etc.). Andererseits kostet die Tür-zu-Tür-Befragung viel Geld [21]. Und dass als Stichprobeneinheit eher die Häuser

als die Menschen erfasst werden, ist ein Problem, das wir bereits bei der Zufallsziffernanwahl kennen gelernt haben. Die mangelnde Teilnahmebereitschaft kann eine ziemliche Herausforderung darstellen. In einem Bericht benötigte man für die erfolgreiche Identifizierung einer Kontrolle im Durchschnitt neun Kontaktversuche [20]; unserer Erfahrung nach werden zuweilen sogar Werte von 150:1 erreicht. Bei Gebäuden, die aus mehreren Wohneinheiten bestehen, müssen zuvor sämtliche Einheiten und die entsprechenden Zugangsmöglichkeiten eruiert werden. Diesem Problem begegnet man aber nicht nur in städtischen Gebieten. In einer Fall-Kontroll-Studie, an der wir beteiligt waren, bekamen es die Interviewer mit Deutschen Schäferhunden und Stacheldrahtzäunen zu tun und wurden sogar von der argwöhnischen Ortspolizei festgenommen [22].

Krankenhauskontrollen

Krankenhauskontrollen sind in Fall-Kontroll-Studien schon häufig eingesetzt und ebenso häufig auch kritisiert worden. Ihre Anwendung erscheint aus mehreren Gründen verlockend: Sie sind leicht erreichbar, ihre Identifizierung und Befragung verursacht nur geringe Kosten, die Informationsqualität ist mit der von Fallpatienten vergleichbar, hohe Teilnahmemotivation und vergleichbares Verhalten bei der Inanspruchnahme medizinischer Leistungen [15]. Doch auch die Nachteile sind nicht unerheblich. Die Verwendung von Krankenhauskontrollen setzt voraus, dass sie für die Hintergrundrate der Exposition innerhalb der Studiengruppe, aus der die Fälle bestehen, repräsentativ sind; die Exposition darf also nicht mit der Erkrankung in Zusammenhang stehen, die zur Hospitalisierung des Kontrollpatienten geführt hat. Am besten lässt sich dieser Fallstrick vermeiden, wenn man diejenigen Patienten als Kontrollen ausschließt, deren Einweisungsdiagnose wahrscheinlich mit der zu untersuchenden Erkrankung in Verbindung steht. So wurden beispielsweise in einer krankenhausbasierten Fall-Kontroll-Studie über Antikonzeption und systemischen Lupus erythema-

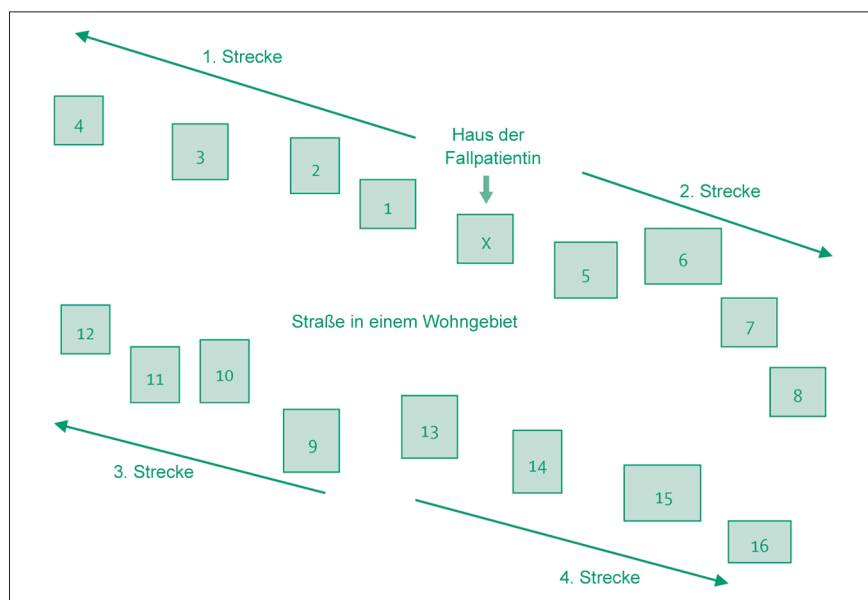


Abb. 4. Nachbarschaftskontrollen. Nach Befragung der Fallpatientin zuhause geht der Untersucher nach einem zuvor festgelegten H-förmigen Schema mit maximal 16 Häusern von Haus zu Haus, bis er drei Kontrollpersonen gefunden und befragt hat. Jedes Viereck steht für ein Haus in derselben Straße.

todes die in die geburtshilfliche und gynäkologische Abteilung aufgenommenen Kontrollpatientinnen aus der Untersuchung ausgeschlossen [23]. Begründet wurde dieser Ausschluss damit, dass Frauen, die in Krankenhäusern der Tertiärversorgung reproduktionsmedizinisch behandelt werden, eine andere geburtshilfliche und gynäkologische Vorgeschichte haben als die meisten anderen Frauen in der Bevölkerung. Für unterschiedliche Krankheiten kann es in einem Krankenhaus unterschiedliche Einzugsgebiete geben; die Kontrollen sollten demselben Einzugsgebiet entstammen wie die Fälle.

Auch durch die Einweisungsrate bedingte systematische Fehler können Probleme verursachen [4, 24]. Wenn z. B. Frauen mit einem Intrauterinpressar (IUP) eine höhere Wahrscheinlichkeit für eine stationäre Behandlung wegen Salpingitis aufweisen als Frauen ohne IUP, dann würde sich dieser Unterschied in einer übertrieben hohen Odds-Ratio für die mit IUP-Anwendung assoziierten Salpingitiden niederschlagen [25].

Verschiedene Berichte lassen darauf schließen, dass Krankenhauskontrollen für die Studiengruppe nicht unbedingt repräsentativ sein müssen, da sie den Fällen eventuell ähnlicher sind als etwa Bevölkerungskontrollen [26, 27]; und in wieder anderen Berichten wurde festgestellt, dass zwischen Krankenhaus- und Bevölkerungskontrollen mitunter erhebliche Unterschiede hinsichtlich Körpergewicht, Rauchverhalten und Krankheitslast bestehen, die sich auf die Wahrscheinlichkeit einer Krankenhaus-einweisung auswirken können [28].

Freunde oder Kollegen als Kontrollen

Manchmal dienen auch die Freunde oder Kollegen der Fallpatienten als Kontrollen – ein Vorgehen, das sowohl Befürworter als auch Kritiker findet. Ein Vorteil ist die Bildung einer Kontrollgruppe, die den Fällen in verschiedenen wichtigen Aspekten wie etwa im sozioökonomischen Status und Bildungsstand ähnlich ist. Fallpatienten zu bitten, potenzielle Kontrollen zu benennen, ist jedoch das genaue Gegenteil von Zufallsauswahl. Die genannten Per-

sonen könnten vielleicht geselliger und umgänglicher sein als andere potenzielle Kontrollen. Das hieße, dass diese Kontrollen nicht repräsentativ wären [15]. Was Randpopulationen betrifft, in denen sozial unakzeptable Verhaltensweisen wie z. B. Drogenabusus untersucht werden sollen, ist man andererseits der Meinung, dass Freundkontrollen praktisch sind und die Wahrscheinlichkeit eines Selektionsbias eher gering ist. In einer Studie wurden Drogenabhängige gebeten, einen Freund zu benennen, der ebenfalls drogenabhängig war (ein neuer Fall) und einen weiteren Freund, der noch nie mit Drogen in Kontakt gekommen war (eine Kontrolle). Aufgrund dieses Kettenverweis- oder Schneeballsystems wurde geschlossen, dass Fälle und Kontrollen derselben Grundgesamtheit entstammten [29].

Angehörige als Kontrollen

Fallpatienten und ihre Angehörigen haben viele Merkmale gemeinsam. Wenn genetische Faktoren als Störgrößen angesehen wurden, hat man zur Kontrolle dieses Bias schon Angehörige eingesetzt [15]. Auch bezüglich zahlreicher anderer Expositionen bestehen Ähnlichkeiten – so geht man beispielsweise davon aus, dass Geschwister auch Gemeinsamkeiten hinsichtlich Ernährung, Umgebung, Lebensstil und sozioökonomischem Status aufweisen. Dienen z. B. Geschwister als Kontrollen, kann der potenzielle Effekt der Familiengröße nicht untersucht werden [15]. Verschiedene Forscher sind zu dem Schluss gekommen, dass die Verwendung von Angehörigen als Kontrollen die Ergebnisse nicht verzerrt, solange die expositionsspezifischen Risiken im Laufe der Zeit stabil bleiben [30].

Wie viele Kontrollgruppen?

Einige Autoren befürworten die Verwendung von zwei getrennten Kontrollgruppen. Bei einheitlichen Ergebnissen, so die Argumentation, steigt dadurch die Glaubwürdigkeit der Befunde [14, 31]. In einer Fall-Kontroll-Studie über Östrogentherapie und Endometriumkarzinom wurden bei-

spielsweise sowohl Krankenhaus- als auch Bevölkerungskontrollen verwendet [32]. Ergibt sich jedoch die missliche Lage, dass die erzielten Ergebnisse unvereinbar sind, dann stellt sich die Frage, welche Resultate ignoriert werden sollen [15]. Ein weiterer sich unmittelbar ergebender Nachteil ist der zusätzliche Zeit- und Ressourcenaufwand. In der oben erwähnten Fall-Kontroll-Studie über das Endometriumkarzinom etwa [32] erhöhte sich durch die Hinzunahme einer Kontrollgruppe aus der Bevölkerung die Anzahl der zu befragenden Studienteilnehmerinnen von 480 auf 801, was einen Anstieg um 67% entspricht. In der Regel raten wir dazu, die bestmögliche Kontrollgruppe auszuwählen [33].

Wie viele Kontrollen pro Fallpatient?

Zuweilen ist der Leser überrascht, wenn er feststellt, dass zwischen der Anzahl der Fälle und Kontrollen in einer Fall-Kontroll-Studie große Unterschiede bestehen; Ärzte erwarten intuitiv ähnliche Gruppengrößen [34]. In einer solchen Ungleichheit spiegelt sich der Versuch der Untersucher wider, die Fähigkeit der Studie zum Nachweis von relevanten Unterschieden – falls tatsächlich vorhanden – zu erhöhen. In Fall-Kontroll-Studien ohne Matching sind annähernd gleiche Zahlen bei Fällen und Kontrollen bei ähnlichen Kosten für Fälle und Kontrollen ausgesprochen effizient. Manchmal ist die Anzahl der Fälle aber nur gering und lässt sich nicht steigern, so etwa bei den im Laufe von 5 Jahren bei jungen Frauen im Bezirk Los Angeles aufgetretenen 11 Fällen eines Leberkarzinoms [35]. Bei einer ebenso kleinen Anzahl von Kontrollen bestünde kaum die Möglichkeit, Assoziationen nachzuweisen. Die Erhöhung der Anzahl von Kontrollen bis zu einem Verhältnis von 4:1 verbessert die Power der Studie. Diese Steigerung ist allerdings nicht linear. Oberhalb des Verhältnisses von ca. 4:1 [36] lässt sich die Studienpower durch die Erhöhung der Anzahl der Kontrollen nur noch geringfügig steigern [37]. Eine Erhöhung des Verhältnisses von Kontrollen zu Fällen wirkt sich auf das Konfi-

denzintervall (die Genauigkeit der Ergebnisse), nicht aber auf die Biasproblematik aus.

Worauf Sie bei Kontrollen achten sollten

Die Validität von Fall-Kontroll-Studien hängt von der Auswahl geeigneter Kontrollgruppen ab. Die Selektion von Kontrollen klingt trügerisch einfach, kann sich jedoch als tückisch erweisen. Kontrollen sollten die Hintergrundhäufigkeit der Exposition in der Grundgesamtheit widerspiegeln und den Fällen deshalb auch in allen wichtigen Aspekten mit Ausnahme der untersuchten Krankheit ähnlich sein. Die Selektion der Kontrollen muss unabhängig von der Exposition erfolgen.

Wenn die Studiengruppe der potenziellen Kontrollen bekannt ist, hat es sich bewährt, alle zu nehmen oder, falls dies nicht durchführbar ist, eine Zufallsstichprobe aus dieser Gruppe zu ziehen. Ist die Gruppe der potenziellen Kontrollen dagegen unbekannt, wird die Auswahl schwierig. Generell bevorzugen wir Personen aus demselben Zeitraum und Ort wie die Fallpatienten.

Suchen Sie nach *einer* guten Kontrollgruppe. Wenn sich die Untersucher hinsichtlich der Eignung einer Kontrollgruppe nicht sicher sind, bilden sie manchmal eine zweite Kontrollgruppe. Bei nur wenigen Fallpatienten führt eine bis zu viermal höhere Anzahl von Kontrollen zur Verbesserung der Studienpower. Die Validität der Untersuchung wird durch eine solche Strategie jedoch nicht gesteigert.

In der Regel führt die Verwendung ungeeigneter Kontrollgruppen nicht nur zu falschen Schlussfolgerungen, sondern kann medizinisch sogar Schaden anrichten [13, 25, 38, 39]. Die Leser von Fall-Kontroll-Berichten sollten daher die Charakteristika der Kontrollen sorgfältig unter die Lupe nehmen. Die Ergebnisse solcher Studien stehen auf Messers Schneide.

Offenlegung von Interessenkonflikten

Hiermit erklären wir, dass keinerlei Interessenkonflikte vorliegen.

Danksagung

Wir danken Willard Cates und David L. Sackett für ihre hilfreichen Anmerkungen zu einer früheren Version des vorliegenden Artikels.

Literatur

- [1] Grimes DA. Clinical research in ancient Babylon: methodologic insights from the book of Daniel. *Obstet Gynecol* 1995;86:1031–4.
- [2] Feinstein AR. *Clinical epidemiology: the architecture of clinical research*. Philadelphia: WB Saunders; 1985.
- [3] Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet* 2002;359:431–4.
- [4] Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;32:51–63.
- [5] Marbach JJ, Schwartz S, Link BG. The control group conundrum in chronic pain case/control studies. *Clin J Pain* 1992;8:39–43.
- [6] Rothman KJ. *Modern epidemiology*. Boston: Little, Brown and Company; 1986.
- [7] Schlesselman JJ. *Case-control studies: design conduct analysis*. New York: Oxford University Press; 1982.
- [8] Chan AT. Aspirin, non-steroidal anti-inflammatory drugs and colorectal neoplasia: future challenges in chemoprevention. *Cancer Causes Control* 2003; 14:413–8.
- [9] Baron JA. Epidemiology of non-steroidal anti-inflammatory drugs and cancer. *Prog Exp Tumor Res* 2003;37: 1–24.
- [10] Zondervan KT, Cardon LR, Kennedy SH. What makes a good case-control study? Design issues for complex traits such as endometriosis. *Hum Reprod* 2002;17:1415–23.
- [11] Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies, I: principles. *Am J Epidemiol* 1991;135:1019–28.
- [12] Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet* 2002;359:341–5.
- [13] Moss AR, Osmond D, Bacchetti P, Chermann JC, Barre-Sinoussi F, Carlson J. Risk factors for AIDS and HIV seropositivity in homosexual men. *Am J Epidemiol* 1987;125:1035–47.
- [14] Perillo MG. Choice of controls in case-control studies. *J Manipulative Physiol Ther* 1993;6:578–85.
- [15] Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls

- in case-control studies, II: types of controls. *Am J Epidemiol* 1992;135:1029–41.
- [16] Hartge P, Brinton LA, Rosenthal JF, Cahill JI, Hoover RN, Waksberg J. Random digit dialing in selecting a population-based control group. *Am J Epidemiol* 1984;120:825–33.
- [17] Olson SH, Kelsey JL, Pearson TA, Levin B. Evaluation of random digit dialing as a method of control selection in case-control studies. *Am J Epidemiol* 1992;135:210–22.
- [18] Voigt LF, Koepsell TD, Daling JR. Characteristics of telephone survey respondents according to willingness to participate. *Am J Epidemiol* 2003;157:66–73.
- [19] Funkhouser E, Macaluso M, Wang X. Alternative strategies for selecting population controls: comparison of random digit dialing and targeted telephone calls. *Ann Epidemiol* 2000;10:59–67.
- [20] Vernick LJ, Vernick SL, Kuller LH (1984) Selection of neighbourhood controls: logistics and fieldwork. *J Chronic Dis* 37: 177–82
- [21] Ryu JE, Thompson CJ, Crouse JR III (1989) Selection of neighbourhood controls for a study of coronary artery disease. *Am J Epidemiol* 129: 407–14
- [22] Rooks JB, Ory HW, Ishak KG, et al. Epidemiology of hepatocellular adenoma: the rate of oral contraceptive use. *JAMA* 1979;242:644–8.
- [23] Grimes DA, LeBolt SA, Grimes KR, Wingo PA. Systemic lupus erythematosus and reproductive function: a case-control study. *Am J Obstet Gynecol* 1985;153:179–86.
- [24] Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248–52.
- [25] Grimes DA. Intrauterine devices and pelvic inflammatory disease: recent developments. *Contraception* 1987;36:97–109.
- [26] West DW, Schuman KL, Lyon JL, Robison LM, Allred R. Differences in risk estimations from a hospital and a population-based case-control study. *Int J Epidemiol* 1984;13:235–9.
- [27] Infante-Rivard C. Hospital or population controls for case-control studies of severe childhood diseases? *Am J Epidemiol* 2003;157:176–82.
- [28] Olson SH, Kelsey JL, Pearson TA, Levin B. Characteristics of a hypothetical group of hospital controls for a case-control study. *Am J Epidemiol* 1994;139:302–11.
- [29] Lopes CS, Rodrigues LC, Sichieri R. The lack of selection bias in a snowball sampled case-control study on drug abuse. *Int J Epidemiol* 1996;25:1267–70.
- [30] Goldstein AM, Hodge SE, Haile RW. Selection bias in case-control studies using relatives as the controls. *Int J Epidemiol* 1989;18:985–9.
- [31] Ibrahim MA, Spitzer WO. The case control study: the problem and the prospect. *Chronic Dis* 1979;32:139–44.
- [32] Hulka BS, Fowler WC Jr, Kaufman DG et al. Estrogen and endometrial cancer: cases and two control groups from North Carolina. *Am J Obstet Gynecol* 1980;137:92–101.
- [33] Moritz DJ, Kelsey JL, Grisso JA. Hospital controls versus community controls: differences in inferences regarding risk factors for hip fracture. *Am J Epidemiol* 1997;145:653–60.
- [34] Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet* 2002;359:966–70.
- [35] Henderson BE, Preston-Martin S, Edmondson HA, Peters R, Pike MC. Hepatocellular carcinoma and oral contraceptives. *Br J Cancer* 1983;48:437–40.
- [36] Kearney DJ, Crump C, Maynard C, Boyko EJ. A case-control study of endoscopy and mortality from adenocarcinoma of the esophagus or gastric cardiac in persons with GERD. *Gastrointest Endosc* 2003;57:823–9.
- [37] Wiebe DJ. Firearms in US homes as a risk factor for unintentional gunshot fatality. *Accid Anal Prev* 2003;35:711–6.
- [38] Rossouw JE. Estrogens for prevention of coronary heart disease: putting the brakes on the bandwagon. *Circulation* 1996;94:2982–5.
- [39] Leventhal JM. Risk factors for child abuse: methodologic standards in case-control studies. *Pediatrics* 1981;68:684–90.

Korrespondenzadresse:

Dr. David Grimes
dgrimes@fhi.org

Anmerkung der Redaktion:

Die Übersetzung dieses Artikels erfolgte durch Frau Karin Beifuss (Stuttgart), die fachliche Bearbeitung übernahm Frau Gerta Rücker (IMBI – Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Freiburg). Beiden sei an dieser Stelle sehr herzlich gedankt.

