

Übersichtsarbeit

Methoden zur Bewertung der Kausalität in Beobachtungsstudien

Teil 27 der Serie zur Bewertung wissenschaftlicher Publikationen

Emilio A. L. Gianicolo, Martin Eichler, Oliver Muensterer, Konstantin Strauch, Maria Blettner

Zusammenfassung

Hintergrund: In der klinisch-medizinischen Forschung wird Kausalität mithilfe der randomisierten kontrollierten Studie (RCT) nachgewiesen. Oft ist aber aus ethischen, manchmal auch aus praktischen Gründen keine RCT möglich und Beobachtungsstudien werden zur Kenntniserlangung herangezogen. Dazu werden zwei Methoden, die bisher in der Medizin wenig genutzt werden, vorgestellt.

Methode: Anhand einer selektiven Literaturrecherche wird eine Einführung in die Methoden, mit der die kausale Inferenz in Beobachtungsstudien bewertet werden, gegeben.

Ergebnisse: Zwei relativ neue Ansätze, Regressions-Diskontinuitäts-Methoden und „interrupted time series“, können unter bestimmten Bedingungen einen kausalen Zusammenhang belegen. Das Regressions-Diskontinuitäts-Design ist ein quasi-experimenteller Ansatz und kann eingesetzt werden, wenn eine kontinuierliche Zuordnungsvariable mit einem Schwellenwert versehen wird. Anhand des Schwellenwerts werden die Patienten in verschiedene Behandlungsschemata eingeordnet. Es wird davon ausgegangen, dass in einem kleinen Intervall um einen Schwellenwert, beispielsweise Cholesterinwerte von 160 mg/dL, Probanden eher zufällig einer von zwei Gruppen zugeordnet werden. Wenn nun oberhalb des Schwellenwertes eine Intervention vorgenommen wird, kann die Gruppe unterhalb als Kontrolle herangezogen werden. Eine spezielle Form des Regressions-Diskontinuitäts-Designs sind die „interrupted time series“. Hier dient die Zeit als Zuordnungsvariable, der sogenannte Cut-off-Punkt. Dies ist oft ein externes Ereignis wie die Einführung eines Rauchverbotes. Im Vorher-Nachher-Vergleich kann festgestellt werden, ob sich der Cut-off-Punkt beziehungsweise das Rauchverbot auf gesundheitliche Parameter wie die Häufigkeit kardiovaskulärer Erkrankungen ausgewirkt hat.

Schlussfolgerung: Mithilfe der beschriebenen Ansätze können Beobachtungsstudien auch kausal interpretiert werden. Vor der Anwendung dieser Methoden sind aber die Voraussetzungen sorgfältig zu prüfen.

Zitierweise

Gianicolo EAL, Eichler M, Muensterer O, Strauch K, Blettner M:
Methods for evaluating causality in observational studies—part 27 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 2020; 117: 101–7.
DOI: 10.3238/arztebl.2020.0101

Institut für Medizinische Biometrie, Epidemiologie und Informatik der Universitätsmedizin Mainz an der Johannes Gutenberg-Universität Mainz: Dr. rer. physiol. Emilio Antonio Luca Gianicolo, Prof. Dr. rer. nat. Konstantin Strauch, Prof. Dr. rer. nat. Maria Blettner

Institute of Clinical Physiology of the Italian National Research Council, Lecce, Italien:
Dr. rer. physiol. Emilio Antonio Luca Gianicolo

Technische Universität Dresden, Medizinische Fakultät Carl Gustav Carus, Medizinische Klinik und Poliklinik I, Dresden: Dr. phil. Martin Eichler

Klinik und Poliklinik für Kinderchirurgie der Universitätsmedizin Mainz an der Johannes Gutenberg-Universität Mainz: Prof. Dr. med. Oliver Muensterer

Institut für Genetische Epidemiologie, Helmholtz Zentrum München – Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg; Lehrstuhl für Genetische Epidemiologie, Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, Ludwig-Maximilians-Universität, München: Prof. Dr. rer. nat. Konstantin Strauch

Das Korrelation nicht Kausalität bedeutet, war oft im Zusammenhang mit der Diskussion um die Auswirkungen der Dieseleexposition im Jahr 2019 zu lesen (1, 2). Diese Binsenweisheit dürfte wohl den meisten klar sein. Schwieriger gestaltet sich die Frage, wie Kausalität eindeutig definiert und nachgewiesen wird (*Kasten 1*). Nach dem Philosophen David Hume ist Kausalität unter zwei Bedingungen gegeben: 1) B folgt immer auf A (oder A ist eine hinlängliche Ursache; „sufficient cause“). 2) Falls A nicht auftritt, tritt B nicht auf (A ist die notwendige Ursache; „necessary cause“) (3). Diese strengen logischen Bedingungen sind in der Medizin nur selten erfüllt. Für die Dieselelaborproblematik würde es bedeuten, dass die Exposition gegenüber Feinstaub immer zu Lungenkrebs führt und Lungenkrebs ohne eine Exposition gegenüber Feinstaub nicht entsteht. Beide Aussagen sind natürlich falsch. Was also ist biologische, medizinische oder epidemiologische Kausalität? In der Medizin wird oft von einem probabilistischen Kausalitätsverständnis ausgegangen. Das heißt, die Exposition gegenüber einem Risikofaktor wie Rauchen oder Dieselelaborgasen erhöht die Wahrscheinlichkeit für die Krankheit, zum Beispiel Lungenkrebs. Dies gilt für viele Therapieansätze. Beispielsweise erhöht eine bestimmte Chemotherapie die Wahrscheinlichkeit des Überlebens nach einer Krebsdiagnose, garantiert sie aber nicht.

In vielen Naturwissenschaften muss die Kausalität in einem Experiment nachgewiesen werden. In der klinisch-medizinischen Forschung dient dazu die randomisierte kontrollierte Studie („randomised controlled trial“ [RCT]) (4). Oft ist aber aus ethischen oder praktischen Gründen keine RCT möglich. Bei der Untersuchung von Risikofaktoren wie der Exposition gegenüber Dieselelaborgasen kann keine zufällige Aufteilung auf Exponierte und Nichtexponierte erfolgen. Darüber hinaus ist eine Randomisierung nicht möglich, wenn untersucht werden soll, ob ein – mit einer Exposition verbundener – Unfall wie der Reaktorunfall in Tschernobyl die Inzidenz oder Mortalität von Krankheiten reduziert oder erhöht. Ähnlich verhält es sich bei der Einführung eines neuen Gesetzes, beispielsweise zum Rauchverbot.

Sind Experimente nicht möglich, wird auf die Beobachtung zurückgegriffen. Hier wird der Gegenstand der Untersuchung, also die mögliche Ursache, nicht ge-

KASTEN 1

Kausalität in epidemiologischen Beobachtungsstudien nach Parascondola (34) et al.

1. Kausalität als Produktion: E führt K herbei. Kausalität ist zu unterscheiden von einer bloßen zeitlichen Abfolge. Es ist nicht hinreichend, dass K immer auf E folgt, K muss E darüber hinaus auf eine Art hervorbringen. Unklar bleibt dabei allerdings, was ein Herbeiführen, Herstellen oder Produzieren bedeutet. Praktisch wird das Produktionsverständnis in der medizinischen Forschung in Grafiken von Wirkmechanismen illustriert.
2. Hinreichende und notwendige Ursachen: Hinreichend ist eine Ursache E, wenn K immer dann geschieht, wenn E eingetreten ist. Notwendig ist sie, wenn K nur dann geschieht, wenn E eingetreten ist. Obwohl dies logisch klar und einfach erscheint, findet sich ein solches Verständnis von deterministischer Kausalität kaum in der Forschungswirklichkeit. Demnach ist das Rauchen weder eine notwendige noch eine hinreichende Ursache für Lungenkrebs. Lungenkrebs kann ohne Tabakrauchexposition entstehen (keine notwendige Ursache), auch folgt auf das Tabakrauchen nicht zwangsläufig Lungenkrebs (keine hinreichende Ursache).
3. Hinlängliche Kausalität („sufficient component cause“): Entwickelt als Reaktion auf die Definition von hinreichenden und notwendigen Ursachen. Der Ansatz geht von einer Mehrzahl von Ursachen aus, die erst im Zusammenspiel einen Effekt entstehen lassen, ohne dass die einzelnen Ursachen alleine jeweils hinreichend Bedeutung erlangen. Auch kann es verschiedene Ursachenkomplexe für einzelne Effekte geben.
4. Probabilistische Kausalität: Hier erhöhen Ursachen die Wahrscheinlichkeit (P) des Eintreffens eines Effektes ($P [K/E] > P [K/nicht E]$). Notwendige und hinreichende Ursachen sind in dieser Hinsicht Grenzfälle, bei denen die Wahrscheinlichkeiten $P(K | nicht E)$ und $P(K | E)$ 0 beziehungsweise 1 sind. Diese Wahrscheinlichkeiten sind im Gegensatz zu den zuerst genannten Ansätzen nicht deterministisch zu werten. Sie können daher Werte zwischen 0 und 1 annehmen. Es wird nicht angenommen, dass ein Effekt eintreffen muss. Diese Sichtweise entspricht dem methodischen Vorgehen in statistisch orientierten Wissenschaften.
5. Kausale Inferenz: Kausales Schließen bezeichnet den Akt der Feststellung einer kausalen Beziehung zwischen verschiedenen Tatbeständen. Dazu werden die Veränderungen, die am Effekt auftreten, analysiert, wenn sich eine Wandlung an der Ursache ergibt. Das kausale Schließen geht im Anspruch über die Feststellung bloßer Assoziationen hinaus und ist mit verschiedenen Konzepten und Schlagwörtern verbunden. In letzter Zeit werden vor allem „counterfactuals“, „potential outcomes“, Kausaldiagramme und Strukturgleichungsmodelle viel diskutiert (36, 37).
6. Triangulation: Nicht alle Fragestellungen lassen sich in Form von Experimenten wie RCT beantworten. Eine Alternative ist der Rückgriff auf einen Methodenpluralismus, wofür sich der Begriff der Triangulation etabliert hat: Das Vertrauen in ein Ergebnis steigt, wenn verschiedene Daten, wissenschaftliche Bereiche, Theorien und Methoden zum selben Ergebnis kommen (35).
7. Konsequenzkriterium: Die Feststellung eines kausalen Zusammenhangs ist mit gesellschaftlichen Konsequenzen (Handeln/Nichthandeln) verbunden. Olsen fordert die Formulierung eines Kriteriums, wann ein solches Handeln erfolgen sollte und wann nicht (7).

zielt und kontrolliert verändert: Es wird protokolliert, wie diese Einflüsse eine Zielvariable, zum Beispiel eine bestimmte Krankheit, verändern.

Die Epidemiologie hat eine Reihe von Überlegungen angestellt, wann auch ohne Experiment von Kausalität gesprochen werden kann, angefangen bei der klassischen Arbeit von Bradford Hill und die von ihm empfohlenen neun Gesichtspunkte für Kausalität (*Kasten 2*) (5) bis hin zu aktuelleren Arbeiten (6, 7).

Neben der statistischen Unsicherheit, die immer entsteht, wenn statt der Gesamtheit aller Betroffenen nur eine Stichprobe analysiert wird (8), sind die Haupthindernisse bei der Untersuchung eines kausalen Zusammenhangs die Störvariablen beziehungsweise Confounder. Diese werden so bezeichnet, weil sie Effekte verschleiern oder vortäuschen können (9). Das Alter ist zum Beispiel ein Confounder, wenn die Assoziation zwischen beruflicher Strahlenexposition und Katarakt untersucht wird (10), denn genauso wie das Kataraktisiko steigt auch die kumulative Strahlenexposition mit zunehmendem Alter.

Für die Analyse von Daten mit bekannten Confoundern existieren verschiedene statistische Methoden, die in anderen Arbeiten (9, 11, 12) bereits vorgestellt wur-

den. Im vorliegenden Artikel gehen wir auf einige neuere Ansätze ein, die bisher in Medizin und Epidemiologie noch wenig Anwendung finden.

Methoden zur Bewertung der kausalen Inferenz aus Beobachtungsstudien

Der Hauptvorteil der RCT besteht darin, dass durch die Randomisierung die Beobachtungseinheiten (Patienten) zufällig auf Behandlungsgruppen aufgeteilt werden. Mögliche bekannte, aber auch unbekannte Confounder sind dadurch in beiden Gruppen zufällig verteilt – auch wenn sich bei der Zufallszuordnung Unterschiede zwischen den Gruppen, die durch Stichprobenvarianz bedingt sind, ergeben können. Wenn eine Randomisierung nicht erfolgen kann, muss der störende Effekt des Confounders bei der Studienplanung oder -auswertung und auch bei der Interpretation beachtet werden.

Klassische Verfahren zur Berücksichtigung von Confoundern bei der Planung der Studie sind die Stratifizierung und das „matched“ Design (13, 14) sowie das sogenannte Propensity-Score-Matching (PSM) (11).

Die bekannteste und am häufigsten verwendete Methode bei der Auswertung der Daten ist die Regres-

sionsanalyse, beispielsweise lineare, logistische oder Cox-Regression (15). Dabei ist der Ausgangspunkt ein mathematisches Modell, das die Wahrscheinlichkeit für das Auftreten des Outcomes durch die bekannten Confounder und den zu untersuchenden Effekt erklären soll.

Die Regressionsanalysen werden bei der Auswertung klinischer oder epidemiologischer Daten angewendet und sind in jeder gängigen statistischen Software vorhanden. Allerdings werden Regressionsverfahren in der Praxis häufig falsch angewandt, weil ihre notwendigen Voraussetzungen nicht überprüft werden. Regressionsanalysen sollten zum Beispiel nicht durchgeführt werden, wenn der Stichprobenumfang zu klein beziehungsweise die Zahl der Variablen zu groß ist oder die Korrelation zwischen den Modellvariablen die Interpretation verhindert (16).

Regressions-Diskontinuitäts-Methoden

Regressions-Diskontinuitäts-Methoden sind in der medizinischen Forschung bisher wenig genutzte Ansätze, um mithilfe von Beobachtungsdaten Ursache und Effekt zu untersuchen (17). Das Regressions-Diskontinuitäts-Design ist ein quasi-experimenteller Ansatz (*Kasten 3*), der in den 1960-er Jahren in der pädagogischen Psychologie entwickelt wurde (18). Es kann eingesetzt werden, wenn eine kontinuierliche Variable („assignment variable“, Zuordnungsvariable) mit einem Schwellenwert, mit dessen Hilfe Patienten in verschiedene Behandlungsschemata eingeordnet werden, versehen wird (*Kasten 4*).

Als Zuordnungsvariable kann zum Beispiel der Cholesterin-Wert im Blut dienen. Personen ab einem Cholesterin-Wert von 160 mg/dL wird eine Therapie verordnet. Liegt der Cholesterin-Wert in der Nähe des Schwellenwertes (160 mg/dL), wird die Person eher zufällig in die eine oder andere Gruppe eingeteilt, da der Cholesterin-Wert (Zuordnungsvariable) mit einem zufälligen Messfehler behaftet ist. In einem kleinen Intervall um den Schwellenwert erfolgt die Aufteilung auf die beiden Gruppen demnach zufällig (18). Die Wirkung der Therapie wird bei Personen in einem engen Bereich um den Schwellenwert ermittelt. Dabei wird davon ausgegangen, dass der Wert mit Messfehlern behaftet ist und sich Personen mit Untersuchungsergebnissen leicht unterhalb oder oberhalb des Schwellenwertes praktisch nicht unterscheiden. Die Zuordnung der Therapie in diesem engen Bereich kann als quasi-zufällig angesehen werden.

Voraussetzungen für die Anwendung dieses Verfahrens sind:

- Die Zuordnungsvariable ist eine stetige Variable, die vor der Behandlung gemessen wurde. Wenn die Zuordnungsvariable komplett unabhängig vom Outcome wäre und keine biologische, medizinische oder epidemiologische Bedeutung hat, ist das Verfahren theoretisch äquivalent zu einer RCT (19).
- Die Behandlung darf nicht die Zuordnungsvariable beeinflussen (18).

KASTEN 2

Bradford-Hill-Kriterien für Kausalität (5)

1. Stärke: Je stärker die beobachtete Assoziation zwischen zwei Variablen, desto unwahrscheinlicher ist, dass die Assoziation ein Ergebnis zufallsbedingter Streuung ist.
2. Konsistenz: Die Assoziation wurde in verschiedenen Studien, Risikopopulationen, Orten, Zeiten und von unterschiedlichen Wissenschaftlern beobachtet.
3. Spezifität: Ein starkes Argument für Kausalität ist es, wenn eine spezifische Bevölkerung an einer spezifischen Krankheit leidet.
4. Zeitliche Beziehung: Die Wirkung muss der Ursache zeitlich nachfolgen.
5. Biologischer Gradient: Die Assoziation zeigt eine Dosis-Wirkungs-Beziehung. Zum Beispiel steigt die Inzidenz von Lungenkrebs mit der Anzahl der täglich gerauchten Zigaretten.
6. Plausibilität: Ein plausibler Mechanismus zwischen Ursache und Wirkung ist hilfreich, aber nicht notwendig. Was heute als plausibel erscheint, hängt vom derzeitigen biologischen Wissen ab.
7. Kohärenz/Stimmigkeit: Die kausale Interpretation der Daten soll nicht im Gegensatz zum biologischen Wissen über eine bestimmte Krankheit stehen.
8. Experiment: Wenn möglich, wird auf experimentelle Evidenz zurückgegriffen.
9. Analogie: Die Assoziation spricht für Kausalität, wenn schon ähnliche Effekte einer ähnlichen Wirksubstanz dokumentiert wurden.

- Es ist nachzuweisen, ob die Baseline-Eigenschaften der Patienten beider Behandlungsgruppen, deren Werte nahe der Schwelle lagen, ähnlich sind, das heißt die Kovariablen inklusive möglicher Confounder. Diese Bilanzierung kann sowohl mit statistischen Verfahren als auch grafisch geprüft werden (20).
- Die optimale Bandbreite der Zuordnungsvariable um die Schwelle muss festgelegt werden: Einerseits sollte sie groß genug sein, um hinreichend große Stichprobenumfänge in den Gruppen zu erhalten. Andererseits sollte sie klein genug sein, damit der Einfluss der Zuordnungsvariable nicht den zu untersuchenden Outcome überdeckt. Methoden dazu werden in der Literatur vorgestellt (21, 22).
- Die Behandlung kann nur auf der Basis der Zuordnungsvariable entschieden werden (deterministische Regressions-Diskontinuitäts-Methoden) oder auf der Basis anderer klinischer Faktoren (Fuzzy-Regressions-Diskontinuitäts-Methoden).

KASTEN 3

Der Wortschatz der Experimente (18)

- **Experiment**
Eine Studie, bei der die Intervention absichtlich eingeführt wird, um eine Wirkung zu beobachten.
- **Randomisiertes Experiment**
Ein Experiment, bei der die Personen, Patienten oder andere Beobachtungseinheiten zufällig einer Therapie- beziehungsweise Interventionsgruppe zugeordnet werden.
- **Quasi-Experiment**
Ein Experiment, bei dem die Beobachtungseinheiten nicht zufällig zugeordnet werden.
- **Naturalexperiment**
Eine Studie, in der ein natürliches Ereignis, zum Beispiel ein Erdbeben, mit einem Vergleichsszenario kontrastiert wird.
- **Nichtexperimentelle Beobachtungsstudie**
Eine Studie, in der Größe und Richtung der Assoziation zwischen Variablen bestimmt werden.

KASTEN 4

Regressions-Diskontinuitäts-Methoden

In dem einfachsten Fall, dass die Regression linear ist, sind die Parameter des folgenden Modells zu schätzen:

$$y_i = \beta_0 + \beta_1 z_i + \beta_2 (x_i - x_c) + e_i$$

Dabei gilt:

- i von 1 bis N sind die statistischen Einheiten
- y ist das Outcome
- β_0 ist der y-Achsenabschnitt („intercept“)
- z ist eine dichotomische Variable (0, 1), die angibt, ob der Patient behandelt (1) oder nicht behandelt (0) wurde
- x ist die Zuordnungsvariable
- x_c ist der Schwellenwert
- β_1 ist der Effekt der Behandlung
- β_2 ist der Regressionskoeffizient der Zuordnungsvariable
- e ist die zufällige Störgröße

KASTEN 5

„Interrupted time series“

Im einfachsten Fall einer „interrupted time series“-Studie folgt die Analyse der Zeitreihe einer stückweisen Regression. Das folgende Modell wird eingesetzt, um sowohl einen Steigungseffekt als auch eine Änderung im Niveau eines Outcomes vor und nach einer Intervention, zum Beispiel der Einführung eines Gesetzes für Rauchverbot, zu untersuchen (*Grafik 2*):

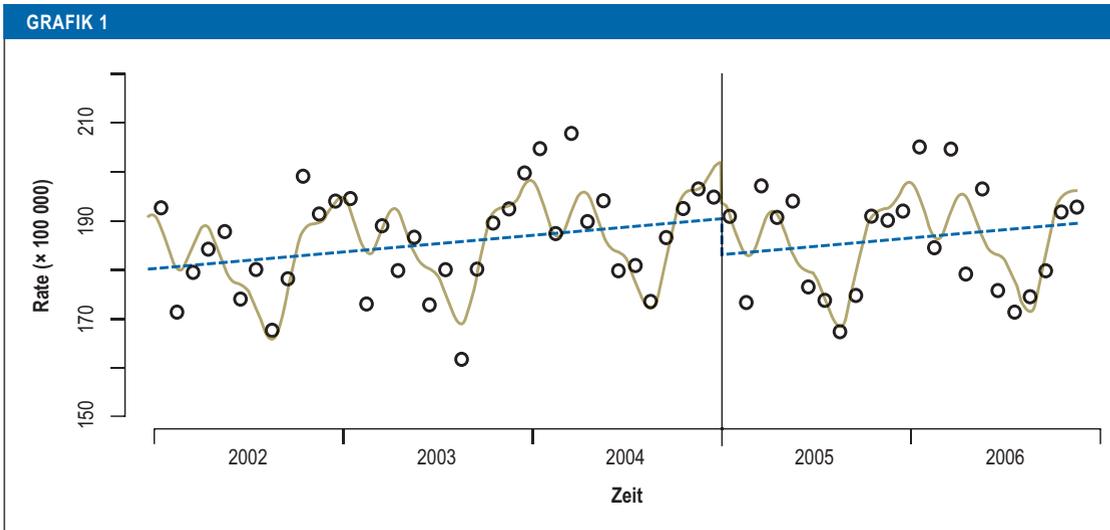
$$y = \beta_0 + \beta_1 \times \text{Zeit} + \beta_2 \times \text{Intervention} + \beta_3 \times \text{Zeit} \times \text{Intervention} + e$$

Dabei gilt:

- y ist das Outcome, zum Beispiel kardiovaskuläre Erkrankungen
- Intervention:
Intervention ist eine Dummy-Variable für die Zeiträume vor (0) und nach der Intervention (1) (zum Beispiel Rauchverbot)
- Zeit ist die Zeit seit dem Beginn der Studie
- β_0 ist die Baseline-Inzidenz von kardiovaskulären Erkrankungen
- β_1 ist die zeitliche Änderung (Steigung) im Niveau der Inzidenz von kardiovaskulären Erkrankungen vor der Einführung des Rauchverbots
- β_2 ist die Änderung im Niveau der Inzidenz von kardiovaskulären Erkrankungen nach der Einführung des Rauchverbots (Niveau-Effekt)
- β_3 ist die Änderung der Zeitabhängigkeit (siehe β_1) nach der Einführung des Rauchverbots (Steigungseffekt)
- e ist die zufällige Störgröße

Beispiel 1: Untersucht werden soll die Ein-Jahres-Mortalität bei Neugeborenen in Abhängigkeit von der Pflegeintensität. Die Pflegeintensität hängt von einem Schwellenwert beim Geburtsgewicht ab. Kinder mit sehr geringem Geburtsgewicht (< 1 500 g) (Gruppe A) erhalten mehr Pflege als schwerere Kinder (Gruppe B) (23). Es soll ermittelt werden, ob sich durch den zusätz-

lichen Pflegeaufwand in Gruppe A die Mortalitätsrate in beiden Gruppen unterscheidet. Angenommen wird, dass Kinder, deren Gewicht sich um den Schwellenwert bewegt, in allen anderen Eigenschaften gleich sind und dass die Zuordnung zu Gruppe A oder B quasi-zufällig erfolgt, da der Messwert mit einem relativ kleinen Fehler behaftet ist. Verglichen werden zum Beispiel Kinder



Die altersstandardisierten stationären Aufnahmeraten in Krankenhäusern von Personen unter 70 Jahren für akute koronare Ereignisse (AKE) vor und nach der Implementation eines Rauchverbots in öffentlichen Räumen in Italien wurden mit entsprechenden Methoden untersucht (30). Beobachtete (Kreise) und vorhergesagte standardisierte Raten (durchgezogene Linien) für Personen unter 70 Jahren. Die gestrichelten Linien zeigen den saisonbereinigten Trend der AKE vor und nach der Einführung des nationalen Rauchverbots.

mit einem Geburtsgewicht von 1 450–1 500 g mit Kindern, deren Geburtsgewicht bei 1 501–1 550 g lag. Auf diese Weise kann untersucht werden, ob die intensivere Pflege die Mortalität beeinflusst.

In diesem Beispiel wird angenommen, dass das Gewicht als Variable mit einem zufälligen Messfehler behaftet ist und dass Patienten mit einem Gewicht nahe der Zuordnungsschwelle zufällig der einen oder der anderen Kategorie zugeordnet werden können. Da das Gewicht aber ein entscheidender Faktor für die Säuglingssterblichkeit ist und Neugeborene mit niedrigem Gewicht eine höhere Kindersterblichkeit als die etwas schwereren Kinder aufweisen (23), darf die Bandbreite nicht zu breit gewählt werden. In dieser Studie weisen die Neugeborenen mit einem Gewicht leicht unterhalb des Schwellenwerts und einer damit verbundenen intensiveren Pflege eine niedrigere Kindersterblichkeit auf als die Kinder, deren Gewicht leicht oberhalb des Schwellenwerts lag und die keine intensivere Pflege erhielten.

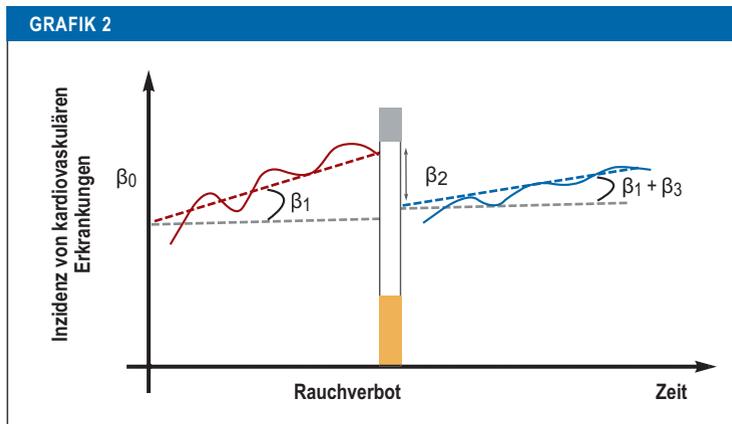
Beispiel 2: Durch ein Regressions-Diskontinuitäts-Studiendesign wurde eine Maßnahme der kanadischen Regierung, ein Mindestalter für Alkoholkonsum von 19 Jahren einzuführen, evaluiert. Die Forscher verglichen die Anzahl von Störungen durch Alkohol, alkoholbedingte Übergriffe, Unfälle und Suizid wenige Monate vor (Gruppe A) und nach dem 19. Geburtstag (Gruppe B). Die Personen in Gruppe B wiesen eine erhöhte Anzahl von alkoholbedingten stationären Behandlungen und Notaufnahmen auf als die Personen in Gruppe A. Mit diesem quasi-experimentellen Ansatz zeigten die Forscher, dass die Maßnahme erfolgreich war (24). Es kann angenommen werden, dass sich die zwei Gruppen nur in Bezug auf das Alter, aber nicht in Bezug auf weitere Eigenschaften, die den Alkoholkonsum beeinflussen, unterscheiden.

„Interrupted time series“

Eine Form von Regressions-Diskontinuitäts-Design sind die „interrupted time series“. Dabei dient die Zeit als Zuordnungsvariable. Der sogenannte Cut-off-Punkt ist oft ein externes Ereignis, das eindeutig zu einem bestimmten Zeitpunkt identifiziert werden kann, zum Beispiel ein Industrieunfall oder eine Gesetzesänderung. Es handelt sich also um einen Vorher-Nachher-Vergleich, bei dem aber vorhandene Zeittrends oder saisonale Schwankungen in der Analyse berücksichtigt werden (*Kasten 5*).

Einige Voraussetzungen für die Verwendung dieses Verfahrens sind zu berücksichtigen (18, 25):

- „Interrupted time series“ sind nur gültig, wenn nur eine einzige Intervention in der Periode vorliegt.
- Die Zeit vor der Intervention muss deutlich von der Zeit nach der Intervention unterschieden werden können.
- Es liegen keine festen Regeln vor, um die Anzahl der benötigten Datenpunkte zu bestimmen. Allerdings sind Studien mit wenigen Zeitpunkten oder kleinen Effektgrößen vorsichtig zu interpretieren. Die Power ist am größten, wenn die Anzahl der Datenpunkte vor und nach der Intervention gleich ist (26).
- Obwohl die Gleichung in *Kasten 5* eine lineare Spezifikation hat, können auch polynomische und andere nichtlineare Regressionsmodelle verwendet werden. Eine sorgfältige Untersuchung der Zeitreihe für nichtlineare Muster ist entscheidend.
- Korreliert eine Beobachtung zum Zeitpunkt *t*, zum Beispiel monatliche Inzidenz von kardiovaskulären Erkrankungen, mit vorausgehenden Beobachtungen (Autoregression), müssen entsprechende statistische Verfahren („auto-regressive integrated moving average“ [ARIMA]-Modelle) eingesetzt werden.



Auswirkung eines Rauchverbots auf die Inzidenz von kardiovaskulären Erkrankungen

Beispiel 1: In einer Studie wurden akute stationäre Aufnahme­raten von kardiovaskulären Erkrankungen vor und nach der Schließung des Heathrow Flughafens aufgrund vulkanischer Asche untersucht, um die Effekte des Fluglärms zu erforschen (27). Die natürliche Intervention fand vom 15. bis 20. April 2010 statt. Die Hospitalisierungen sanken in der Bevölkerungsgruppe, die in dem Stadtteil mit dem höchsten Fluglärm wohnte. Allerdings war die Zahl der Beobachtungspunkte zu gering, um eine klare Evidenz für den Zusammenhang zu postulieren.

Beispiel 2: In einer Studie wurden stationäre Aufnahme­raten in Krankenhäuser vor und nach der Implementa­tion des Rauchverbots (Intervention) in öffentlichen Räumen in Italien untersucht (28). Die Intervention fand im Januar 2004 (Cut-off) statt. Gemessen wurde die Anzahl stationärer Krankenhausaufenthalte aufgrund akuter koronarer Ereignisse von Januar 2002 bis November 2006 (Grafik 1). In der Analyse wird die Saisonabhängigkeit berücksichtigt und zusätzlich eine Effektmodifikation für die Altersgruppen < 70 Jahren und ≥ 70 Jahren ermittelt. Die Aufnahme­raten verringerten sich für Personen unter 70 Jahren.

Diskussion

Die notwendige Unterscheidung zwischen Kausalität und Korrelation wird zwar in wissenschaftlichen Diskussionen oft betont, leider aber auch oft nicht stringent beachtet. In der Medizin und Epidemiologie muss man zudem meistens von einer probabilistischen Kausalität ausgehen, also einer Erhöhung der Wahrscheinlichkeit, dass das untersuchte Ereignis eintritt. Beispielhaft sei hier die Strahlenforschung genannt, bei der strikt zwischen sogenannten deterministischen und probabilistischen beziehungsweise stochastischen Strahlenschäden unterschieden wird (29). Deterministische Strahlenschäden treten bei einer bestimmten, meist hohen Strahlendosis mit Sicherheit auf (Verbrennungen bis Tod). Hingegen ist die Erhöhung der Krebsmortalität nach Strahlenexposition ein stochastischer Strahlenschaden. Epidemiologische Beobachtungen und biologische Experimente sollten ge-

meinsam evaluiert werden, um die probabilistische Kausalität zu untermauern (Kasten 1).

Auch wenn die RCT als Goldstandard in der klinischen Forschung nicht an Bedeutung verloren hat, ist sie nicht immer durchführbar. Deshalb sind Erkenntnisse aus Beobachtungsstudien unverzichtbar. Für diese sind schon bei der Planung Vorkehrungen zu treffen, um Störfaktoren zu eliminieren oder zu berücksichtigen. Zusätzlich ist eine sorgfältige Analyse der Daten notwendig. Dennoch ist eine einzelne Beobachtungsstudie praktisch nie in der Lage, einen kausalen Zusammenhang zu bestimmen.

In diesem Artikel wurden zwei Verfahren präsentiert, die wegen ihrer relativen Einfachheit viel Potenzial für einen breiteren Einsatz in der Medizin und der Epidemiologie haben können (30). Beide Verfahren sollten aber nur nach sorgfältiger Prüfung der Voraussetzungen angewandt werden. In Regressions-Diskontinuitäts-Methoden muss die Annahme der Kontinuität geprüft werden. Das heißt, es ist zu prüfen, ob andere Eigenschaften in der Behandlungs- und Kontrollgruppe gleich beziehungsweise bilanziert sind. Zusätzlich müssen die Regel der Entscheidung und der Schwellenwert der kontinuierlichen Zuordnungsvariable bekannt sein. Die Regressions-Diskontinuitäts-Methoden können kausale Schlussfolgerungen generieren. Inwieweit diese Schlussfolgerungen verallgemeinerbar sind ist jedoch begrenzt, wenn die Behandlungseffekte über den Bereich der Zuordnungsvariable heterogen sind. Die Schätzung der Effektgröße gilt nur lokal über einen beliebig kleinen Bereich um den Schwellenwert. Zudem muss geprüft werden, ob eine lineare Beziehung zwischen Outcome und Zuordnungsvariable besteht und ob eine Interaktion zwischen der Behandlungs- sowie der Zuordnungsvariable zu berücksichtigen ist.

Auch in der „interrupted time series“-Analyse muss die Annahme der Kontinuität geprüft werden. Außerdem ist das Verfahren nur dann valide, wenn man ausschließen kann, dass zu dem gleichen Zeitpunkt keine anderen Interventionen stattfanden (20). Schließlich muss die Form der Zeitreihe berücksichtigt werden und gegebenenfalls müssen komplexere statistische Methoden angewandt werden, um beispielsweise das Phänomen der Autoregression zu berücksichtigen.

Oft deuten Ergebnisse aus Beobachtungsstudien auf Zusammenhänge, die dann in weiteren Studien und Experimenten untermauert oder verworfen werden, hin. Die Erkenntnisse der Strahlenforschung beruhten zunächst im Wesentlichen auf Beobachtungen aus Hiroshima und Nagasaki (31). Diese wurden durch epidemiologische Studien in anderen strahlenexponierten Populationen, beispielsweise beruflicher oder medizinischer Exposition, durch physikalische Betrachtungen sowie durch biologische Experimente unterstützt (32). Als klassisches Beispiel soll die Beobachtung von Snow angeführt werden (33): Vor dessen Beobachtungen waren die biologischen Ursachen von Cholera nicht bekannt. Snow beobachtete, dass die Verschmutzung einiger Brunnen in einem kausalen Zusammenhang mit dem Auftreten der Cholera stehen müsste. Daraufhin wurden schließlich die hygienischen Zustände verbessert und der Infektion mit

dem Cholera-Erreger vorgebeugt. Fälle wie diese zeigen, dass es sinnvoll sein kann, bereits aus den Beobachtungen Konsequenzen zu ziehen (6). Darüber hinaus demonstrieren die genannten Fälle, dass weitere Untersuchungen erfolgen müssen, um die Kausalität zu bestätigen.

Interessenkonflikt

Die Autoren erklären, dass kein Interessenkonflikt besteht.

Manuskriptdaten

eingereicht: 2. 8. 2019, revidierte Fassung angenommen: 18. 11. 2019

Literatur

1. Köhler D: Feinstaub und Stickstoffdioxid (NO₂): Eine kritische Bewertung der aktuellen Risikodiskussion. Dtsch Arztebl 2018; 115(38): A-1645.
2. Deutsche Gesellschaft für Epidemiologie, Deutsche Gesellschaft für Medizinische Informatik Biometrie und Epidemiologie, Deutsche Gesellschaft für Public Health, Deutsche Gesellschaft für Sozialmedizin und Prävention: Offener Brief bzw. Stellungnahme auf den Webseiten der beteiligten Fachgesellschaften 2019. www.dgepi.de/assets/News/84b5207b3d/NOxFeinstaubStellungnahme2019_01_29.pdf (last accessed on 11 January 2020).
3. Hume D: An enquiry concerning human understanding. LaSalle: Open Court Press 1784.
4. Lorenz E, Köpke S, Pfaff H, Blettner M: Cluster-randomized studies—part 25 of a series on evaluating scientific publications. Dtsch Arztebl Int 2018; 115: 163–8.
5. Hill AB: The environment and disease: association or causation? Proc R Soc Med 1965; 58: 295–300.
6. Dekkers OM: The long and winding road to causality. Eur J Epidemiol 2019; 34: 533–5.
7. Olsen J, Jensen UJ: Causal criteria: time has come for a revision. Eur J Epidemiol 2019; 34: 537–41.
8. du Prel JB, Hommel G, Röhrig B, Blettner M: Confidence interval or p-value? Part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106: 335–9.
9. Hammer GP, du Prel JB, Blettner M: Avoiding bias in observational studies: part 8 in a series of articles on evaluation of scientific publications. Dtsch Arztebl Int. 2009; 106: 664–8.
10. Scheidemann-Wesp U, Gianicolo EAL, Camara RJ, et al.: Ionising radiation and lens opacities in interventional physicians: results of a German pilot study. J Radiol Prot 2019; 39: 1041–59.
11. Kuss O, Blettner M, Borgermann J: Propensity Score: an alternative method of analyzing treatment effects. Part 23 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2016; 113: 597–603.
12. Rensing M, Blettner M, Klug SJ: Data analysis of epidemiological studies: part 11 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107: 187–92.
13. Röhrig B, du Prel JB, Wachtlin D, Blettner M: Types of study in medical research: part 3 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106: 262–8.
14. Röhrig B, du Prel JB, Blettner M: Study design in medical research: part 2 of a series on the evaluation of scientific publications. Dtsch Arztebl Int 2009; 106: 184–9.
15. Schneider A, Hommel G, Blettner M: Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2010; 107: 776–82.
16. Hartung J, Elpelt B, Klöesener KH: Statistik – Lehr- und Handbuch der angewandten Statistik. München: Oldenbourg 2005. 204–14.
17. Thistlewaite DL, Campbell DT: Regression-discontinuity analysis: an alternative to the ex-post facto experiment. J Educ Psychol 1960; 51: 309–17.
18. Shadish W, Cook T, Campbell D: Experimental and quasi-experimental designs for generalized causal inference. Belmont, USA: Wadsworth Cengage Learning™ 2002.
19. Lee DS, Lemieux T: Regression discontinuity designs in economics. J Econ Lit 2010; 48: 281–355.
20. Barnighausen T, Oldenburg C, Tugwell P, et al.: Quasi-experimental study designs series-paper 7: assessing the assumptions. J Clin Epidemiol 2017; 89: 53–66.
21. Moscoe E, Bor J, Barnighausen T: Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. J Clin Epidemiol 2015; 68: 122–33.
22. Oldenburg CE, Moscoe E, Barnighausen T: Regression discontinuity for causal effect estimation in epidemiology. Curr Epidemiol Rep 2016; 3: 233–41.
23. Almond D, Doyle JJ, Kowalski AE, Williams H: Estimating marginal returns to medical care: evidence from at-risk newborns. Q J Econ 2010; 125: 591–634.
24. Callaghan RC, Sanches M, Gately JM, Cunningham JK: Effects of the minimum legal drinking age on alcohol-related health service use in hospital settings in Ontario: a regression-discontinuity approach. Am J Public Health 2013; 103: 2284–91.
25. Bernal JL, Cummins S, Gasparrini A: Interrupted time series regression for the evaluation of public health interventions: a tutorial. Int J Epidemiol 2017; 46: 348–55.

Kernaussagen

- Unter Berücksichtigung verschiedener Bedingungen kann aus Beobachtungsstudien eine kausale Beziehung abgeleitet werden.
- Störvariablen (Confounder) sind ein Haupthindernis beim Nachweis eines kausalen Zusammenhangs, da diese einen Effekt verschleiern oder vortäuschen können.
- Durch eine randomisierte Zuordnung werden bekannte und unbekannte Störvariablen auf die zu vergleichenden Gruppen gleichmäßig verteilt.
- Bei der Regressions-Diskontinuitäts-Methode wird davon ausgegangen, dass in einem engen Intervall um einen Schwellenwert die Aufteilung in beide Gruppen zufällig stattfindet und somit auch die Störgrößen zufällig verteilt sind.
- Die „interrupted time series“ ist eine Variante der Regressions-Diskontinuitäts-Methode, wobei durch einen bestimmten Zeitpunkt eine Vorher- und Nachhergruppe festgelegt und damit die Störgrößen zufällig verteilt werden.

26. Zhang F, Wagner AK, Ross-Degnan D: Simulation-based power calculation for designing interrupted time series analyses of health policy interventions. J Clin Epidemiol 2011; 64: 1252–61.
27. Pearson T, Campbell MJ, Maheswaran R: Acute effects of aircraft noise on cardiovascular admissions – an interrupted time-series analysis of a six-day closure of London Heathrow Airport caused by volcanic ash. Spat Spatiotemporal Epidemiol 2016; 18: 38–43.
28. Barone-Adesi F, Gasparrini A, Vizzini L, Merletti F, Richiardi L: Effects of Italian smoking regulation on rates of hospital admission for acute coronary events: a country-wide study. PLoS One 2011; 6: e17419.
29. International Commission on Radiological Protection: Recommendations of the ICRP – ICRP Publication 26. Oxford: Pergamon Press 1977 (last accessed on 17 Januar 2020).
30. Bor J, Moscoe E, Mutevedzi P, Newell ML, Barnighausen T: Regression discontinuity designs in epidemiology: causal inference without randomized trials. Epidemiology 2014; 25: 729–37.
31. Preston DL, Kusumi S, Tomonaga M, et al.: Cancer incidence in atomic bomb survivors. Part III. Leukemia, lymphoma and multiple myeloma, 1950–1987. Radiat Res 1994; 137 (2 Suppl): 68–97.
32. United Nations Scientific Committee on the Effects of Atomic Radiation UNSCEAR: Sources, effects and risks of ionizing radiation. Report to the general assembly, with scientific annexes 2016.
33. Snow J: Cholera and the water supply in the south districts of London in 1854. J Public Health Sanit Rev 1856; 2: 239–57.
34. Parascandola M, Weed DL: Causation in epidemiology. J Epidemiol Community Health 2001; 55: 905–12.
35. Munafò MR, Davey Smith G: Repeating experiments is not enough. Verifying results requires disparate lines of evidence – a technique called triangulation. Nature 2018; 553: 399–401.
36. Hernán MA, Robins JM: Causal inference: what if. Boca Raton: Chapman & Hall/ CRC 2020.
37. Pearl J, Mackenzie D: The book of why. The new science of cause and effect. New York: Penguin 2018.

Anschrift für die Verfasser

Dr. rer. physiol. Emilio Antonio Luca Gianicolo
 Institut für Medizinische Biometrie, Epidemiologie und Informatik
 Universitätsmedizin der Johannes Gutenberg-Universität Mainz
 Abteilung Epidemiologie und Versorgungsforschung
 Obere Zahlbacher Str. 69, 55131 Mainz
 emilio.gianicolo@uni-mainz.de

Zitierweise

Gianicolo EAL, Eichler M, Muensterer O, Strauch K, Blettner M: Methods for evaluating causality in observational studies—part 27 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2020; 117: 101–7. DOI: 10.3238/arztebl.2020.0101

► Die englische Version des Artikels ist online abrufbar unter: www.aerzteblatt-international.de