

# Was wollen und was können wir durch empirische Kausalforschung erfahren?<sup>1</sup>

Rolf Steyer

## Zusammenfassung

Anhand zweier Beispiele wird zunächst gezeigt, dass und warum die Analyse von Mittelwerten und ihren Unterschieden zwischen Gruppen zu kausalen Fehlschlüssen führen kann. Danach werden die Grundbegriffe der Theorie individueller und durchschnittlicher kausaler Effekte eingeführt. Anhand eines Beispiels zur nonorthogonalen Varianzanalyse wird dann demonstriert, dass schon diese elementaren Begriffe genügen, um ein altes, bisher nicht gelöstes Problem der Angewandten Statistik zu lösen. Schließlich wird die Theorie um den Begriff der Unkonfundiertheit ergänzt, der insbesondere – aber nicht nur – für die kausale Analyse von nichtexperimentellen Daten relevant ist.

## 1. Einleitung

Empirische Kausalforschung ist in der Psychologie und ihren Nachbardisziplinen allgegenwärtig. Psychologen wollen z. B. wissen, ob und wie ein spezielles Therapieverfahren einem anderen bei einer bestimmten Störungsform überlegen ist oder ob eine bestimmte Unterrichtsform gegenüber einer anderen zu besseren Resultaten hinsichtlich eines bestimmten Lernkriteriums führt. Politiker interessiert, ob eine sozialpolitische Maßnahme zur gewünschten Reduktion bestimmter Gewaltescheinungen beiträgt, ob Einsparungen im Haushalt des Bundes oder ob eine bestimmte Steuerreform zur Verminderung des Defizits beitragen. Sportwissenschaftler fragen, ob ein bestimmtes Training in einer bestimmten Sportart zu höheren Leistungen führt als ein anderes Training. Umweltwissenschaftler untersuchen, ob von Menschen hervorgerufener Kohlendioxidausstoß zur Erderwärmung beiträgt. Mediziner wollen wissen, ob Rauchen zu Lungenkrebs, fettreiche Nahrung und mangelnde Bewegung zu Herzinfarkt führen etc. Ohne kausale Theorien über die Wirkungen unseres Handelns wären wir handlungsunfähig. Handeln – im Gegensatz zu Verhalten – setzt Absicht und diese zumindest eine implizite Theorie oder Erwartung über den Effekt des Handelns voraus. Auch im Bereich des menschlichen Lernens gewinnt die Sichtweise des *kausalen Lernens* gegenüber der des reinen asso-

---

<sup>1</sup> Überarbeitete Fassung meiner Antrittsvorlesung an der Friedrich-Schiller-Universität Jena, am 19. November 1997.

ziativen Lernens zunehmend an Bedeutung (s. z. B. Waldmann, 2000, 2001).

*Empirische Kausalforschung* ist überall dort angesagt, wo wir unsicher über die Wirkung bestimmter Handlungsalternativen sind. Die o. g. und andere Beispiele zeigen: Korrekte kausale Theorien sind lebens-, wenn nicht sogar überlebenswichtig für einen Staat, für soziale Gruppen und für einzelne Individuen, im privaten wie im professionellen Handeln.<sup>2</sup>

Entsprechend seiner Bedeutsamkeit hat das Thema „Kausalität“ viele Wissenschaftsphilosophen, Statistiker, Methodiker und Einzelwissenschaftler angezogen. Auch wenn es an dieser Stelle nicht möglich ist, einen Überblick über die verschiedenen Kausalitätstheorien zu geben, der diesen Namen verdienen würde (s. dazu eher Steyer, 1992, Kap. 2), sollen doch einige Autoren und die zugehörigen Stichworte in Erinnerung gerufen werden.

Wie es kaum anders sein könnte, hat Aristoteles erste wichtige Unterscheidungen eingeführt. Auch Humes (1739/1888) Charakterisierung einer kausalen Beziehung durch „Kontiguität“, „zeitliche Aufeinanderfolge“ und „konstante Verbindung“ ist als Markstein in der Ideengeschichte des Kausalitätsbegriffs zu nennen. Noch einschlägiger sind wohl die Millschen (1843/1862) Methoden der wissenschaftlichen Forschung, insbesondere die *Unterschiedsmethode*: die Variation der möglichen Ursache bei Konstanthaltung aller anderen Einflussvariablen und Feststellung der Kovariation der Ursache mit der zu erklärenden Variablen. Hier sind wir schon ganz dicht an der heute üblichen experimentellen Vorgehensweise.

Die moderne nichtexperimentelle Kausalforschung basiert im wesentlichen auf dem Prinzip der *statistischen Konstanthaltung* der Störvariablen. Dieses Prinzip wurde in einer Vielzahl von Arbeiten im Rahmen der Literatur zur Pfadanalyse (z. B. des Genetikers Wright, 1918, 1921, 1923, 1934, 1960a, 1960b) und zu Strukturgleichungsmodellen (Bollen, 1989; Jöreskog, 1993) ausgearbeitet und verfeinert. Darüber hinaus wurde dort die Unterscheidung zwischen direkten, indirekten und totalen Effekten eingeführt (s. dazu weitere Angaben in Steyer, 1992, Kap. 2). Seit einigen Jahren wurden auch wichtige Fortschritte im Bereich kausaler Modelle mit qualitativen Variablen erzielt (s. z. B. Pearl, 1993, 1995, 2000).

Nicht unerwähnt bleiben darf in diesem Kontext die in der Erziehungswissenschaft und der Psychologie populäre *Theorie interner Validität* von Campbell und Stanley (1963) bzw. Cook und Campbell (1979, 1986), die im deutschsprachigen Raum nicht zuletzt durch Bredenkamp (1969,

---

<sup>2</sup> Wenn ich hier nicht zögere, von „überlebenswichtig“ zu sprechen, denke ich an die großen politischen Katastrophen, die wir Deutsche und Europäer im vergangenen Jahrhundert erlebt haben.

1980) und später von Hager (1987) kritisiert und weiterentwickelt wurde. Die Kernfrage ist hier die nach alternativen Erklärungen für eine beobachtete Abhängigkeit. Viele fruchtbare und wichtige methodologische Diskussionen zu Fragen des Designs von Untersuchungen und der Datenanalyse basieren auf diesem Begriff.

Unter empirisch arbeitenden Wissenschaftlern weniger einflussreich ist die Vielzahl stochastischer Kausalitätstheorien. Hervorheben möchte ich hier die auf Suppes (1970) zurückgehenden Arbeiten von Stegmüller (1983) und Spohn (1980, 1983, 1990, 1991).

Auch die Arbeiten und Ansätze in der Statistik haben sich bisher noch nicht sehr weit in der empirischen Kausalforschung durchgesetzt. Zu nennen sind hier die Arbeiten von Rubin (1974, 1977, 1985), Rubin und Thayer (1978), Holland (1986, 1988), Rosenbaum (1984a, 1984b, 1984c), Sobel (1994), Little und Yau (1998) sowie Jo (2002), die auf Neyman (1923/1990) zurückgehen. Gleiches gilt auch für die Arbeiten von Pearl (1993, 1995), die auf der Literatur zu graphischen Modellen (z. B. Whittaker, 1990) aufbauen.

Auch wenn ich hier nur auf einen kleinen Teil der Autoren hinweisen konnte, die sich dem Thema „Kausalität“ gewidmet haben, wird doch deutlich, dass dieses Thema auf viele Wissenschaftler eine große Faszination ausübt. Wie kommt diese Faszination zustande? Ziel dieses Beitrags ist, dies ein Stück weit verständlich zu machen und dabei einige elementare Konzepte und Sätze der Kausalitätstheorie zu vermitteln.

## 2. Der Vergleich von Mittelwerten und relativen Häufigkeiten

Was tun wir üblicherweise, wenn wir wissen wollen, ob eine Variable  $X$  einen Effekt auf eine Variable  $Y$  hat? Das Nahe liegende ist zu prüfen, ob die beiden Variablen miteinander *kovariieren*. Dies kann man z. B. mit der Berechnung der Korrelation zwischen diesen beiden Variablen tun. Ist  $X$  eine Variable mit nur wenigen Ausprägungen, dann kann man auch die Mittelwerte von  $Y$  zwischen den verschiedenen Ausprägungen von  $X$  miteinander vergleichen. Kann auch  $Y$  nur zwei Werte annehmen, so werden die relativen Häufigkeiten, mit denen  $Y$  den einen der beiden Werte annimmt, miteinander verglichen. Betrachten wir ein einfaches Beispiel!

-----  
Etwa hier Tabelle 1 einfügen

---

-----

In Tabelle 1 seien die Ergebnisse einer Datensammlung zur Evaluation einer Behandlung zusammengestellt. Das Kriterium sei die Variable „Erfolg“ (z. B. nicht rückfällig zu werden nach einer Alkoholismustherapie) mit den Ausprägungen „Ja“ und „Nein“. Wie würde man über die Fortführung der Behandlung entscheiden, wenn man mit diesen Daten konfrontiert wäre?

Standard ist es, den Anteil der Erfolgreichen bei den Behandelten mit dem entsprechenden Anteil bei den nicht Behandelten zu vergleichen. Bei den Behandelten beträgt der Anteil der Erfolgreichen 50%, bei den nicht Behandelten dagegen 60% (s. Abb. 1). Bei diesem Vergleich kommen wir also zu dem für die Behandlung Urteil, dass sie ihr Geld nicht wert ist.

-----

Etwa hier Abbildung 1 einfügen

-----

## 2.1 Das Simpson-Paradox

Ganz anders allerdings sieht die Antwort aus, wenn man innerhalb der Gruppe der Männer dieselbe Betrachtung anstellt. Hier sind nun 40% erfolgreich bei den Behandelten und nur 30% bei den nicht Behandelten (s. Tab. 2a). Anders als in der Gesamtpopulation scheint die Behandlung bei den Männern also erfolgreich zu sein.

-----

Etwa hier Tabelle 2 einfügen

-----

Wie sieht es nun bei der Gruppe der Frauen aus? Unserer Erwartung zufolge müssten eigentlich die nicht Behandelten in starkem Maße erfolgreicher sein als die Behandelten, um die Ergebnisse in der Gesamtgruppe erklären zu können. Ein Blick in die Tabelle 2b zeigt jedoch, dass uns unser gesunder Menschenverstand hier trügt. Auch bei den Frauen ist der Anteil der Erfolgreichen bei den Behandelten größer als bei den nicht Behandelten.

In Abbildung 2 sind die betreffenden Anteile dargestellt. Wir sehen: in beiden Geschlechtsgruppen ist offenbar die Behandlung erfolgreich, in der Gesamtgruppe, bestehend aus eben diesen Männern und Frauen, jedoch nicht. In anderen Worten: Wir können sowohl Männer als auch

Frauen erfolgreich behandeln, nicht jedoch Menschen. So könnte das paradox erscheinende Fazit der Vergleiche der relativen Häufigkeiten in den Abbildungen 1 und 2 lauten.

-----

Etwa hier Abbildung 2 einfügen

-----

Wie kann das sein? In jeder der beiden Subpopulationen ein positiver Effekt, und in der Gesamtpopulation, die sich genau aus diesen beiden Subpopulationen zusammensetzt, ein negativer Effekt! Müsste nicht in der Gesamtpopulation der Durchschnitt der Effekte in den Subpopulationen zu finden sein? Wieso läuft unser intuitives Denken hier in die Irre? Im Wesentlichen liegt es am gleichzeitigen Auftreten zweier Bedingungen: Erstens, die ungleichen Erfolgswahrscheinlichkeiten bei Männern und Frauen (s. die rechte Spalte der Tab. 2;  $375/1000$  vs.  $725/1000$ ) und zweitens, die ungleiche Behandlungswahrscheinlichkeit ( $750/1000$  bei den Männern,  $250/1000$  bei den Frauen). Kommt beides zusammen, entstehen derartige, paradox erscheinende Phänomene.

## 2.2 Ein weiteres Beispiel

Ein weiteres Beispiel kann uns helfen zu verstehen, wie solche Paradoxa zustande kommen. Daneben dient dieses Beispiel zur Einführung einiger Grundbegriffe der Theorie individueller und durchschnittlicher kausaler Effekte. Bemerkenswert ist dieses Beispiel auch insofern, als *bei jeder einzelnen* Person ein *positiver* Effekt einer Behandlung vorliegt, aber dennoch eine *negative* Mittelwertsdifferenz auftreten kann, wenn man diese Personen in zwei Gruppen, die der Behandelten und die der nicht Behandelten, aufteilt.

-----

Etwa hier Tabelle 3 einfügen

-----

Machen wir ein Gedankenexperiment! In Tabelle 3a sei eine Population von acht Personen mit ihren Erwartungswerten in einer Experimental- und einer Kontrollbedingung, sowie deren Differenz dargestellt. Natürlich sind in jedem normalen Anwendungsfall diese Werte unbekannt, aber wir können uns vorstellen, dass die angegebenen individuellen Erwartungswerte das empirische Phänomen genauso steuern, wie die Wahrscheinlichkeit die möglichen Ergebnisse bei einem Würfelwurf. Würden wir diese individuellen Erwartungswerte kennen, so würden wir den kausalen Effekt der

Behandlung  $E$  gegenüber der Behandlung  $K$  für jede einzelne Person, d. h. die *individuellen kausalen Effekte* kennen – der Traum jedes empirischen Kausalforschers und jedes angewandten Psychologen: Tue ich  $E$ , erwarte ich bei der Person  $u_1$  den Wert 82 auf der betrachteten Lebenszufriedenheitsskala, tue ich  $K$  erwarte ich dagegen nur den Wert 68. Also tue ich besser  $E$ . Bei der Person  $u_6$  erwarte ich bei  $E$  den Wert 131, bei  $K$  nur den Wert 119. Eine idealere Situation können wir uns kaum vorstellen!

Leider kann eine Person in der Regel entweder nur der Experimental- oder nur der Kontrollbedingung unterworfen werden, nicht aber beiden Bedingungen zugleich. Entweder ich therapiere nach Verfahren  $A$  oder nach Verfahren  $B$ , entweder ich unterrichte nach Lehrmethode  $A$  oder nach Lehrmethode  $B$ , etc. Paul Holland (1986) hat dies das *fundamentale Problem der kausalen Inferenz* genannt. Es hat zur Folge, dass wir in der Regel die individuellen kausalen Effekte auch nicht schätzen können.

Allerdings kann man unter bestimmten Bedingungen etwas über den Durchschnitt der individuellen kausalen Effekte, den *durchschnittlichen kausalen Effekt* (s. die letzte Zahl rechts unten in der Tab. 3a) in einer Population und in Subpopulationen erfahren. Wären wir in der Lage, die Spaltenmittelwerte der ersten beiden Spalten in Tabelle 3a zu schätzen, indem wir *einige Personen unter der einen ( $E$ ) und andere Personen unter der anderen experimentellen Bedingung ( $K$ ) beobachten*, so hätten wir mit der Differenz dieser Spaltenmittelwerte eine Schätzung für den durchschnittlichen kausalen Effekt (s. die letzte Zeile von Tab. 3a).

Die Spaltenmittelwerte der ersten beiden Spalten der Tabelle 3a zu schätzen, indem wir einige Personen unter der einen ( $E$ ) und andere Personen unter der anderen experimentellen Bedingung ( $K$ ) beobachten, kann man auf unterschiedliche, mehr oder weniger kluge Weisen versuchen. In Tabelle 3b ist eine weniger kluge Weise dargestellt, deren Betrachtung aber dennoch lehrreich ist. Dabei wurden die ersten vier Personen der Experimentalbedingung zugeordnet, die letzten vier der Kontrollbedingung (s. die fett markierten Zahlen). In der Experimentalbedingung ergibt sich dabei ein Mittelwert der vier ausgewählten Werte von 95, in der Kontrollbedingung von 125. Die Differenz beträgt  $-30$ , die miserabelste Schätzung für den durchschnittlichen kausalen Effekt ( $+ 10$ ), die in diesem Beispiel auftreten kann.

Was ist hier passiert? Offenbar ist die Zuweisung der Personen zu den beiden Bedingungen,  $E$  und  $K$ , nicht unabhängig von den Erwartungswerten der Personen in den beiden experimentellen

Bedingungen (s. die zweite Zahlenspalte in Tab. 3a).<sup>3</sup> Im Gegenteil, die Personen mit den niedrigsten Erwartungswerten wurden der Experimental-, die mit den höchsten Erwartungswerten der Kontrollbedingung zugeordnet. Die Konsequenz ist für die kausale Interpretierbarkeit der resultierenden Mittelwertsdifferenz verheerend.

Etwas besser, aber immer noch gründlich daneben liegt die Berechnung des durchschnittlichen kausalen Effekts durch die Mittelwertsdifferenz bei der in Tabelle 3c dargestellten Zuordnung der Personen zu den beiden Bedingungen  $E$  und  $K$ . Statt +10 erhält man die Mittelwertsdifferenz -10.

Erst bei der in Tabelle 3d dargestellten Zuordnung kommt man zum richtigen Ergebnis, d. h. erst bei einer Auswahl, in der eine ausgewogene Berücksichtigung von Personen mit hohen und niedrigen Erwartungswerten in beiden Bedingungen realisiert wird, ist die resultierende Mittelwertsdifferenz unverfälscht in dem Sinn, dass sie gleich dem durchschnittlichen kausalen Effekt ist.

Wie lassen sich diese Überlegungen auf reale Experimente übertragen, in denen die Erwartungswerte der Personen nicht bekannt sind? Die klassische und wohlbekannteste Antwort ist, die Personen zufällig auf die experimentellen Bedingungen aufzuteilen. Man spricht in diesem Kontext von *Randomisierung* und vom *randomisierten Experiment*.

Randomisierung garantiert, dass die Erwartungswertdifferenz  $\mu_E - \mu_K$  (im Beispiel der Tab. 4 mit der vorletzten Spalte:  $115 - 105 = 10$ )<sup>4</sup> gleich dem durchschnittlichen kausalen Effekt ist. In Stichproben weichen dann die entsprechenden Mittelwertsdifferenzen nur zufällig vom durchschnittlichen kausalen Effekt ab. Neben der Vergleichbarkeit der erwarteten Outcome-Variablen zwischen den experimentellen Bedingungen stellt man mit diesem Vorgehen sogar eine Vergleichbarkeit der experimentellen Bedingungen hinsichtlich aller möglichen Eigenschaften der Personen her, die diesen Bedingungen nach dem Zufallsprinzip zugewiesen werden.

Hängt die Wahrscheinlichkeit der Zuweisung der Personen dagegen von den erwarteten Outcome-Variablen ab (s. Tab. 4 mit der letzten Spalte; die Zahlen sind ansonsten willkürlich ge-

---

<sup>3</sup> Diese beiden Spalten der individuellen Erwartungswerte, die wir im folgenden als *erwartete Outcome-Variablen* bezeichnen, treten in der hier vorgestellten Version der Theorie der individuellen und durchschnittlichen kausalen Effekte an die Stelle der Rubinschen *potenziellen Outcome-Variablen*  $Y_0$  und  $Y_1$  (s. auch Holland, 1986).

<sup>4</sup> Der Erwartungswert  $\mu_E$  lässt sich nach der allgemeingültigen Formel  $\mu_E = \sum_i \mu_{E,u_i} P(u_i|E)$  berechnen. Die dazu notwendigen Zahlen findet man in Tabelle 4, wobei allerdings die dort angegebenen individuellen Behandlungswahrscheinlichkeiten  $P(E | u_i)$  in die bedingten Wahrscheinlichkeiten  $P(u_i|E)$  umgerechnet werden müssen. Das Entsprechende gilt natürlich für  $\mu_K$ .

wählt), erhält man eine Erwartungswertdifferenz  $\mu_E - \mu_K$  (im Beispiel:  $103,3 - 116,6 = -13,3$ ), die deutlich vom durchschnittlichen kausalen Effekt abweicht.

-----  
 Etwa hier Tabelle 4 einfügen  
 -----

Man kann sich leicht überlegen, dass die zufällige Zuweisung der Personen zu den experimentellen Bedingungen überflüssig ist, wenn die Population homogen ist, d. h. wenn die erwarteten Outcome-Variablen innerhalb jeder experimentellen Bedingung konstant ist, wenn also alle Personen völlig gleich hinsichtlich der betrachteten Variablen funktionieren. In Tabelle 5 ist veranschaulicht, was damit gemeint ist. Funktionieren alle Personen hinsichtlich der betrachteten Variablen  $X$  und  $Y$  völlig gleich, spielt es keine Rolle, wie man sie auf die experimentellen Bedingungen aufteilt. Natürlich ist klar, dass diese Homogenität in unseren empirischen Disziplinen nur selten vorkommt. Man kann allerdings versuchen, sich ihr durch die Auswahl von Subpopulationen anzunähern. Dies entspricht der bereits erwähnten Millschen Idee der Konstanthaltung aller anderen Einflussgrößen.

-----  
 Etwa hier Tabelle 5 einfügen  
 -----

### 3. Die Theorie individueller und durchschnittlicher kausaler Effekte

Fassen wir die bisher eingeführten Begriffe und die damit gewonnenen Erkenntnisse zusammen (s. Abb. 3)! Eine Erwartungswertdifferenz  $\mu_E - \mu_K$  bzgl. einer Kriteriumsvariablen  $Y$  zwischen zwei Bedingungen  $E$  und  $K$ , der *Prima-facie-Effekt* (s. Abb. 3), ist *nicht per se* das, woran wir in der empirischen Kausalforschung interessiert sind.

-----  
 Etwa hier Abbildung 3 einfügen  
 -----

Ein *individueller kausaler Effekt* einer Bedingung  $E$  bzgl. einer Bedingung  $K$  auf eine Kriteriumsvariable  $Y$  für eine Person  $u$  ist die Differenz zwischen dem Erwartungswert von  $Y$  unter der Bedingung  $E$  und dem Erwartungswert von  $Y$  unter der Bedingung  $K$  bei der gleichen Person  $u$ .

Der *durchschnittliche kausale Effekt* einer Bedingung  $E$  bzgl. einer Bedingung  $K$  auf eine



Kriteriumsvariable  $Y$  ist der Durchschnitt der individuellen Erwartungswerte über die betrachtete Population von Personen.

Eine Erwartungswertdifferenz  $\mu_E - \mu_K$  zwischen zwei experimentellen Bedingungen kann unter bestimmten, günstigen Voraussetzungen (s. unten) kausal unverfälscht sein. *Kausale Unverfälschtheit* eines Prima-facie-Effekts  $\mu_E - \mu_K$  heißt: die Differenz  $\mu_E - \mu_K$  ist gleich dem durchschnittlichen kausalen Effekt.

Dies sind die wichtigsten Grundbegriffe. Zwei wichtige Theoreme, die diese Grundbegriffe miteinander in Verbindung bringen, werden in Abbildung 4 illustriert. Diese Theoreme sind letztlich die Grundlage für die Techniken des Experimentierens.

Eine erste hinreichende Bedingung für die kausale Unverfälschtheit des Prima-facie-Effekts  $\mu_E - \mu_K$  ist eine Zuweisung der Personen zu den experimentellen Bedingungen, die *unabhängig* von der Person (und ihren Eigenschaften) ist. Eine zweite hinreichende Bedingung für die kausale Unverfälschtheit des Prima-facie-Effekts ist die Homogenität der Population.

-----

Etwa hier Abbildung 4 einfügen

-----

Diese Ergebnisse erscheinen zu Recht nicht besonders überraschend. Tatsächlich ist bisher lediglich eine Präzisierung des Begriffs der kausalen Unverfälschtheit des Prima-facie-Effekts  $\mu_E - \mu_K$  zu vermerken. Alle anderen Ergebnisse entsprechen ganz unserem intuitiven Vorwissen. Ob sich die Mühe der Präzisierung gelohnt hat, wird sich erst dann zeigen, wenn es gelingt, bisher ungelöste Probleme einer Lösung zu zuführen. Einige solcher Probleme gibt es im Bereich der nonorthogonalen Varianzanalyse und des Designs und der Analyse nichtrandomisierter Untersuchungen. Beginnen wir zunächst mit dem Bekannten, der Analyse orthogonaler zweifaktorieller Versuchspläne!

#### 4. Design und Analyse orthogonaler Versuchspläne

Die oben zusammengefassten Ergebnisse lassen sich direkt auf Fragen des Designs und der Analyse orthogonaler varianzanalytischer Versuchspläne anwenden. In Tabelle 6 ist ein Beispiel für einen solchen orthogonalen varianzanalytischen Versuchsplan dargestellt, bei dem der Faktor A drei experimentelle Bedingungen (hier drei Therapieverfahren) und der Faktor B eine Eigenschaft der Versuchspersonen vor Beginn der Behandlung darstellt (hier die „Behandlungsbedürftigkeit“ in drei Abstuf-

ungen, stark, mittel und schwach behandlungsbedürftig).

-----

Etwa hier Tabelle 6 einfügen

-----

Dabei sei man wie folgt vorgegangen: Es wurden zunächst  $n = 120$  Personen ausgewählt, die am Experiment mitwirken sollten. Nach einer Anamnese wurde festgestellt, wie viele und welche dieser Probanden stark, mittel bzw. schwach behandlungsbedürftig sind. (Im Beispiel seien dies 30, 60 bzw. 30.) Innerhalb jeder dieser drei Bedürftigkeitsgruppen wurden *nach dem Zufallsprinzip* drei gleich große Gruppen gebildet, die jeweils nach einem der drei Therapieverfahren behandelt wurden. Nach einer angemessenen Zeit wurde bei allen Probanden die Kriteriums- oder Outcome-Variable erhoben.

Geht man nach diesem Schema vor, so kann man die Mittelwertsunterschiede zwischen den drei experimentellen Bedingungen, d. h. die Differenzen zwischen den Zeilenmittelwerten, als unverfälschte Schätzungen der durchschnittlichen kausalen Effekte interpretieren. Dies ist die gängige und wohlbegründete Praxis der orthogonalen Varianzanalyse.

Auch innerhalb jeder der drei Bedürftigkeitsgruppen lassen sich die Mittelwertsunterschiede zwischen den drei experimentellen Bedingungen als Schätzungen der durchschnittlichen kausalen Effekte interpretieren, allerdings nicht bezogen auf die Gesamtpopulation, sondern als Durchschnitt bzgl. der jeweiligen Subpopulation.

## 5. Nonorthogonale Designs

Wie sieht es jedoch aus, wenn wir Vorwissen über die Wirkungsweise der drei Therapieformen haben und daher z. B. aus ethischen Gründen nicht bereit sind, bei jeder der drei Ausprägungen der Behandlungsbedürftigkeit jede der drei Therapieformen *mit gleicher Wahrscheinlichkeit* anzuwenden? Nehmen wir an, wir könnten uns zwar zu einer zufälligen Zuweisung der Probanden zu den drei Therapieverfahren entschließen, *aber mit unterschiedlichen Wahrscheinlichkeiten, je nach Ausprägung der Behandlungsbedürftigkeit*. Dieses Verfahren nennen wir *bedingte Randomisierung*. Die Bedürftigen könnten dann eine größere Behandlungswahrscheinlichkeit haben als die Nichtbedürftigen. Die Zuordnungswahrscheinlichkeiten seien dabei proportional zu den in Tabelle 7 angegebenen Anzahlen der Beobachtungen in den Zellen, den *Zellenfrequenzen*. Für eine stark bedürftige Person sei also z. B. die Wahrscheinlichkeit, mit der Therapie  $x_1$  behandelt zu werden, 20/30

und für eine mittel bedürftige Person sei die Wahrscheinlichkeit, mit der Therapie  $x_2$  behandelt zu werden, 40/60. Zerstören diese ungleichen Wahrscheinlichkeiten, den Therapiebedingungen zugeordnet zu werden, die kausale Unverfälschtheit?

-----  
 Etwa hier Tabelle 7 einfügen  
 -----

Die Antwort heißt „jein“. „Ja“, wenn man die normalen Zeilenmittelwerte betrachtet und „nein“, wenn man die Daten richtig auszuwerten weiß. Kausal unverfälscht – im Sinne des durchschnittlichen kausalen Effekts in der entsprechenden Subpopulation – bleiben weiterhin die Mittelwertsunterschiede zwischen den Therapiebedingungen *innerhalb* jeder der drei Ausprägungen des Faktors „Behandlungsbedürftigkeit“. Darüber hinaus kann man hier auch die durchschnittlichen kausalen Effekte in der Gesamtpopulation berechnen, die sich allerdings nicht mehr als Differenz der (normalen) Zeilenmittelwerte ergeben. Stattdessen muss man *aus den Zellenmittelwerten durch Gewichtung mit den relativen Spaltenhäufigkeiten adjustierte Zeilenmittelwerte* berechnen. Letztere sind in diesem Beispiel alle drei gleich, nämlich gleich 100. Demnach sind in diesem Beispiel auch die durchschnittlichen kausalen Effekte in der Gesamtpopulation gleich null, ein Ergebnis, das keines der gängigen Verfahren der nonorthogonalen Varianzanalyse liefert, wie sie in SAS, SPSS, SYSTAT oder ähnlichen Softwarepaketen angeboten werden. Hier dient die Kausalitätstheorie also der Ableitung neuer Analyseverfahren, die alles andere als selbstverständlich und altbekannt sind.

-----  
 Hier etwa Abbildung 5 einfügen  
 -----

In Abbildung 5 sind die o. g. für nonorthogonale Designs relevanten Ergebnisse zusammengefasst. Dabei repräsentiert die Variable Z mit ihren Ausprägungen z den Faktor „Behandlungsbedürftigkeit“. Details findet man in Steyer (2003), Steyer, Nachtigall, Wüthrich-Martone und Kraus (2002) und Wüthrich-Martone (2001).

## 6. Analyse nichtexperimenteller Beobachtungsstudien

Auch aus reinen Beobachtungsstudien, in denen wir die natürlich auftretende Korrelation verschiedener Variablen erfassen, wollen wir oft auf kausale Abhängigkeiten schließen. In epidemiologischen Untersuchungen, in denen es bspw. um den Zusammenhang zwischen Rauchen und Lungenkrebs,

fettreiche Ernährung und Herzinfarkt, Strahlenbelastung in der Nähe von Kernkraftwerken und Leukämie, Stress und Krankheitsanfälligkeit etc. geht, sind kausale Interpretationen von Abhängigkeiten und Mittelwertsunterschieden von zentraler Bedeutung.

Bleiben wir bei den einfachen Fällen, in denen sich Abhängigkeiten als Mittelwertsunterschiede darstellen lassen! Voraussetzungsgemäß liegt bei den in diesem Abschnitt betrachteten Fällen weder ein randomisiertes, noch ein bedingt randomisiertes Experiment vor. Auch Homogenität, die ja ebenfalls eine kausale Unverfälschtheit einer Mittelwertsdifferenz garantieren würde, kann nicht ohne weiteres angenommen werden (s. Abb. 4). Nachdem also in solchen Studien diese bisher behandelten hinreichenden Bedingungen für die kausale Unverfälschtheit von Mittelwertsunterschieden ausfallen, gibt es dann wenigstens notwendige Bedingungen, die aus der Behauptung folgen, dass eine Mittelwertsdifferenz kausal unverfälscht ist?

Die Antwort heißt leider „nein“. Aus der Aussage der kausalen Unverfälschtheit einer Mittelwertsdifferenz lässt sich nichts logisch ableiten, was sich *in der betreffenden Studie* empirisch überprüfen ließe. Das liegt ganz einfach daran, dass die individuellen kausalen Effekte und damit auch deren Durchschnitt über eine Menge von Personen (die Population) nicht schätzbar sind. Damit ist die Behauptung, eine konkrete Mittelwertsdifferenz sei kausal unverfälscht, d. h.  $PFE_{EK} = ACE_{EK}$  (s. Abb. 3), *in der betreffenden Studie* nicht falsifizierbar.

Bevor wir die Konsequenzen dieser Feststellung weiter erörtern, soll sie an einem Beispiel illustriert werden. Angenommen, jemand behauptet, dass der Unterschied zwischen Rauchern und Nichtrauchern in der Wahrscheinlichkeit, an Lungenkrebs zu sterben, im oben definierten Sinn (d. h.  $PFE_{EK} = ACE_{EK}$ ) kausal unverfälscht ist.<sup>5</sup> Die individuellen kausalen Effekte sind hier nicht schätzbar, weil jede Person entweder Raucher oder aber Nichtraucher ist. Damit ist auch deren Durchschnitt über eine Menge von Personen, die Population, nicht schätzbar. Wie sollte man dann also prüfen, ob ein festgestellter Unterschied zwischen Rauchern und Nichtrauchern in der Wahrscheinlichkeit, an Lungenkrebs zu sterben, gleich dem Durchschnitt der individuellen kausalen Effekte ist?

Meine oben getroffene Feststellung, dass eine Aussage, eine konkrete Mittelwertsdifferenz sei kausal unverfälscht, in der betreffenden Studie nicht empirisch falsifizierbar ist, hat wichtige Konsequenzen. Sie sei daher noch einmal wohlbedacht! Im Prinzip könnten wir eine solche Aussage doch in einem *weiteren*, randomisierten Experiment überprüfen. Daher habe ich oben betont: „*in der*

*betreffenden Studie*“. In der Praxis stellen sich diesem Vorgehen allerdings oft Probleme in den Weg. Dort, wo wir uns auf die Analyse von reinen Beobachtungsdaten stützen müssen, lässt sich ja gerade kein randomisiertes Experiment durchführen. Selbst im Dienst der Wissenschaft können wir keine Zufallsstichprobe aus der Population ziehen, die betreffenden Personen zwingen, zu Rauchern zu werden, und den Anteil, der an Lungenkrebs stirbt, mit dem entsprechenden Anteil bei einer Zufallsauswahl von Personen vergleichen, die wir zwingen, Nichtraucher zu bleiben.

Was man jedoch tun kann, ist zu prüfen, ob wir gute Gründe haben anzunehmen, dass eine Mittelwertsdifferenz bzw. eine Differenz in den Wahrscheinlichkeiten kausal unverfälscht ist. Ein guter Grund wäre, dass wir bereit sind anzunehmen, dass eine *hinreichende Bedingung* für die kausale Unverfälschtheit der betreffenden Mittelwertsdifferenz erfüllt ist. Da es neben den bereits erwähnten Bedingungen der Randomisierung und der Homogenität noch weitere hinreichende Bedingungen für die kausale Unverfälschtheit gibt (s. dazu Steyer, Gabler, von Davier und Nachtigall, 2000), ist es hier wichtig, dass man die *schwächste* hinreichende Bedingung für die kausale Unverfälschtheit prüft, die noch einer empirischen Überprüfung zugänglich ist.

Es würde an dieser Stelle zu weit führen, hier diese Bedingung, die *Unkonfundiertheit*, allgemein zu definieren. Dazu sei auf Steyer, Gabler und Rucai (1996) sowie auf Steyer, Gabler, von Davier und Nachtigall (2000) verwiesen. Im wesentlichen ist der Begriff der Unkonfundiertheit dadurch definiert, dass die Erwartungswerte, deren Differenz kausal unverfälscht sein soll, hier also  $\mu_E$  und  $\mu_K$ , selbst immer Erwartungswerte der subpopulationsspezifischen Erwartungswerte sind. Wenn wir zwei experimentelle Bedingungen,  $E$  und  $K$  betrachten, und wenn  $Z$  eine Variable ist, welche die Population in Subpopulationen partitioniert (z. B. das Geschlecht, der Familienstand etc.), dann müssen demnach gelten:

$$\mu_E = \sum_z \mu_{E, Z=z} P(Z=z),$$

$$\mu_K = \sum_z \mu_{K, Z=z} P(Z=z),$$

wobei hier über alle Werte  $z$  der Variablen  $Z$  summiert werden soll. Gilt dies für jede solche Variable  $Z$ , dann definiert dies die *Unkonfundiertheit* der bedingten Erwartungswerte  $\mu_E$  und  $\mu_K$ .<sup>6</sup> Entscheidend in diesem Kontext ist, dass die Unkonfundiertheit sowohl bei Randomisierung als auch bei Ho-

---

<sup>6</sup> Dabei sei angemerkt, dass ein solcher Wahrscheinlichkeitsunterschied auch als Mittelwertsdifferenz aufgefasst werden kann, nämlich als Mittelwertsdifferenz einer Variablen mit den Werten 0 und 1. Dabei steht in diesem Beispiel der Wert 1 für das Ereignis, an Lungenkrebs zu sterben, und der Wert 0 für das Komplementäreignis.

mogenität immer erfüllt ist und ihrerseits die kausale Unverfälschtheit impliziert (s. Abb. 6). Es sei auch angemerkt, dass erst die Unkonfundiertheit garantiert, dass kein Simpson-Paradox mehr auftreten kann. Mit anderen Worten: trotz Gültigkeit der kausalen Unverfälschtheit eines Prima-facie-Effekts kann es passieren, dass in der Gesamtpopulation ein negativer Prima-facie-Effekt vorliegt, obwohl diese in Subpopulationen mit jeweils positiven Prima-facie-Effekten zerlegt werden kann (s. Steyer, Gabler, von Davier, Nachtigall & Buhl, 2000). In diesem Falle würde also nicht die Analyse auf Gesamtpopulations-, sondern die auf Subpopulationsebene verfälschte Effekte liefern. Erst die obige Unkonfundiertheitsbedingung garantiert, dass solche Phänomene nicht mehr auftreten können. (Zur Verallgemeinerung dieser Unkonfundiertheitsbedingung für kontinuierliche Variablen  $X$  und  $Z$  siehe Steyer, 1992).

-----

Hier etwa Abbildung 6 einfügen

-----

## 7. Zusammenfassung und Ausblick

Was wollen und was können wir also in der empirischen Kausalforschung erfahren? Was wir erfahren *wollen*, ist etwas über die *individuellen kausalen Effekte*. Worüber wir unter günstigen Umständen etwas erfahren *können*, ist deren *Durchschnitt* in einer Gesamtpopulation oder in verschiedenen Subpopulationen. Die günstigen Umstände kann man in einem randomisierten Experiment herstellen, in dem man die Personen den experimentellen Bedingungen nach dem Zufallsprinzip zuordnet. Woran wir im Kontext empirischer Kausalforschung nicht *per se* interessiert sind, sind Mittelwerte und deren Differenzen.

Insbesondere die o. g. Erkenntnis über das randomisierte Experiment ist nicht unbedingt neu. Neu ist auch nicht, dass man aus der Theorie der individuellen und durchschnittlichen kausalen Effekte ableiten kann, wie die Daten einer orthogonalen Varianzanalyse analysiert werden sollten. Für die *nonorthogonale Varianzanalyse* haben wir allerdings ein altes Problem lösen können. Ähnliches ist übrigens auch für die *Kovarianzanalyse* zu erwarten, wenn die Kovariate und die experimentelle Variable nicht unabhängig sind.

---

<sup>6</sup> Für eine ausführlichere Darstellung der Theorie, verschiedener Theoreme und Beispiele sei auf Steyer, Gabler, von Davier und Nachtigall (2000) verwiesen.

In solchen Untersuchungen, in denen die Personen den Behandlungsbedingungen *nicht* zufällig zugewiesen werden können, bleibt die kausale Modellierung schwierig. Die Unkonfundiertheitsbedingung ist nicht verifizierbar, aber doch falsifizierbar. Sie ist nicht verifizierbar, weil sie immer nur für *einige* von unendlich vielen potenziellen Störvariablen überprüft werden kann. Sie ist falsifizierbar, weil der Nachweis der Konfundiertheit bzgl. einer einzigen potenziellen Störvariablen Z genügt, um die Unkonfundiertheit bzgl. aller potenziellen Störvariablen zu widerlegen. Die Unkonfundiertheit ist m. W. die schwächste empirisch falsifizierbare hinreichende Bedingung für die kausale Unverfälschtheit eines Prima-facie-Effekts. Scheitern in einer konkreten Anwendung Versuche zu ihrer Falsifikation, so wird die kausale Unverfälschtheit eines Prima-facie-Effekts plausibler. Gelingt die Falsifikation der Unkonfundiertheit für eine spezielle Störvariable, so bleibt die kausale Interpretation der Prima-facie-Effekte *in den Subpopulationen*, die durch die betreffende Störvariable erzeugt werden. Die betreffende Störvariable wäre dann also konstant zu halten.

Vieles wäre noch zu sagen, was bereits erforscht ist, vieles andere bleibt noch zu entdecken. Es gilt, sowohl Designs der Datenerhebung und Methoden der Datenanalyse neu zu entwickeln, als auch bekannte Designs daraufhin zu untersuchen, ob sie sich für kausale Modellierungen eignen. Die theoretischen Grundlagen dazu sind geschaffen und warten auf ihre Weiterentwicklung und kreative Anwendung. Klar ist, dass inhaltliche Forschung, die nicht auf dem gesicherten Fundament einer wohlverstandenen Kausalitätstheorie beruht, Gefahr läuft, in die Irre zu führen und auf die Dauer nicht verantwortbar ist. Hierin liegt eine große Herausforderung und Chance für die Methodenlehre der Sozial- und Verhaltenswissenschaften.

## Literaturverzeichnis

- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bredenkamp, J. (1969). Experiment und Feldexperiment. In C. F. Graumann (Ed.), *Handbuch der Psychologie: Sozialpsychologie, 1. Halbband: Theorien und Methoden* (pp. 332-373). Göttingen: Hogrefe.
- Bredenkamp, J. (1980). *Theorie und Planung psychologischer Experimente*. Darmstadt: Steinkopff.
- Campbell, D. T. & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L.Gage (Ed.), *Handbook on research on teaching* (pp. 171-246). Chicago, IL: Rand McNally.
- Cook, T. D. C. (1986). The causal assumptions of quasi-experimental practice. *Synthese*, 68, 141-180.
- Cook, T. D. C. & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Boston, MA: Houghton Mifflin.
- Hager, W. (1987). Grundlagen einer Versuchsplanung zur Prüfung empirischer Hypothesen in der Psychologie. In G. Lüer (Ed.), *Allgemeine experimentelle Psychologie* (pp. 43-264). Stuttgart: Gustav Fischer.
- Holland, P. (1986). Statistics and causal inference (with comments). *Journal of the American Statistical Association*, 81, 945-970.
- Holland, P. W. (1988). Causal inference, path analysis, and recursive structural equations



models. *Sociological Methodology*, 18, 449-484.

Hume, D. (1888). *A treatise on human nature (Original work published 1739/40)*. Oxford: Clarendon Press.

Jo, B. (2002). Statistical power in randomized intervention studies with noncompliance. *Psychological Methods*, 7, 178-193.

Jöreskog, K. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park: Sage.

Little, R. J. & Yau, L. H. Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods*, 3, 147-159.

Mill, J. S. (1862). *System der deductiven und inductiven Logik. Erster Theil. (Übersetzung von J. Schiel. 2. deutsche, nach der fünften des Originals erweiterte Auflage. Original erschienen 1843)*. Braunschweig: Vieweg.

Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles. Section 9 (reprint 1990). *Statistical Science*, 5, 465-472.

Novick, M. R. (1980). Statistics as psychometrics. *Psychometrika*, 45, 411-424.

Pearl, J. (1993). Comment: Graphical models, causality, and intervention. *Statistical Science*, 8, 266-269.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82, 669-710.

Pearl, J. (2000). *Causality - Models, reasoning, and inference*. Cambridge, MA: Cam-

bridge University Press.

Rosenbaum, P. R. (1984a). Conditional permutation tests and the propensity score in observational studies. *Journal of the American Statistical Association*, 79, 565-574.

Rosenbaum, P. R. (1984b). From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment. *Journal of the American Statistical Association*, 79, 41-48.

Rosenbaum, P. R. (1984c). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society, Series A*, 147, 656-666.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66, 688-701.

Rubin, D. B. (1977). Assignment of treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2, 1-26.

Rubin, D. B. (1985). The use of propensity scores in applied Bayesian inference. *Bayesian Statistics*, 2, 463-472.

Rubin, D. B. & Thayer, D. (1978). Relating tests given to different samples. *Psychometrika*, 43, 3-10.

Sobel, M. E. (1994). Causal inference in latent variable models. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis* (pp. 3-35). Thousand Oaks, CA: Sage.

Spohn, W. (1980). Stochastic independence, causal independence, and shieldability. *Jour-*

*nal of Philosophical Logic*, 9, 73-99.

Spohn, W. (1983). *Deterministic and probabilistic reasons and causes* (Rep. No. 19).

Dordrecht u. Boston: D. Reidel.

Spohn, W. (1990). *Direct and indirect causes* (Rep. No. 9). Dordrecht: Kluwer Academic Publishers.

Spohn, W. (1991). A reason for explanation: Explanations provide stable reasons. In W. Spohn, B. C. van Fraassen, & B. Skyrmes (Eds.), *Existence and explanation* (pp. 165-196). Dordrecht: Kluwer.

Stegmüller, W. (1983). *Erklärung, Begründung, Kausalität (Probleme und Resultate der Wissenschaftstheorie and Analytischen Philosophie, Band I.)*. Berlin: Springer.

Steyer, R. (1992). *Theorie kausaler Regressionsmodelle*. Stuttgart: Gustav Fischer Verlag.

Steyer, R. (2003). *Wahrscheinlichkeit und Regression*. Berlin: Springer.

Steyer, R., Gabler, S. & Rucai, A. A. (1996). Individual causal effects, average causal effects, and unconfoundedness in regression models. In F. Faulbaum & W. Bandilla (Eds.), *Softstat '95. Advances in statistical Software 5* (pp. 203-210). Stuttgart: Lucius & Lucius.

Steyer, R., Gabler, S., von Davier, A. A. & Nachtigall, C. (2000). Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research Online*, 5, 55-86.

Steyer, R., Gabler, S., von Davier, A. A., Nachtigall, C. & Buhl, T. (2000). Causal regression models I: Individual and average causal effects. *Methods of Psychological Research Online*,

5, 39-71.

Steyer, R., Nachtigall, C., Wüthrich-Martone, O. & Kraus, K. (2002). Causal regression models III: Covariates, conditional, and unconditional average causal effects. *Methods of Psychological Research Online*, 7, 41-68.

Suppes, P. (1970). *A probabilistic theory of causality*. Amsterdam: North-Holland.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 53-76.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychometric Bulletin & Review*, 8, 600-608.

Whittaker, J. (1990). *Graphical models in applied multivariable statistics*. Chichester: John Wiley.

Wright, S. (1918). On the nature of size factors. *Genetics*, 3, 367-374.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557-585.

Wright, S. (1923). The theory of path coefficients - A reply to Niles' criticism. *Genetics*, 8, 239-255.

Wright, S. (1934). The method of path coefficients. *Annals of Mathematical Statistics*, 5, 161-215.

Wright, S. (1960a). Path coefficients and path regressions: Alternative or complementary concepts? *Biometrics*, 16, 189-202.

Wright, S. (1960b). The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrics*, 16, 423-445.

Wüthrich-Martone, O. (2001). *Causal modeling in psychology with qualitative independent variables*. Aachen: Shaker.

### **Über den Autor**

Steyer, Rolf, geboren 1950, Diplom im Fach Psychologie 1977 an der Universität Göttingen, Promotion im Fach Psychologie 1982 an der Universität Frankfurt am Main, Habilitation im Fach Psychologie 1989 an der Universität Trier. Von 1994 bis 1995 Wissenschaftlicher Leiter am Zentrum für Methoden, Umfragen und Analysen (ZUMA) in Mannheim, von 1995 bis 1996 Professor für Methodik und Diagnostik der Psychologie an der Universität Magdeburg; seit 1996 Inhaber des Lehrstuhls für Methodenlehre und Evaluationsforschung am Institut für Psychologie der Universität Jena, seit September 2002 Prorektor der Friedrich-Schiller-Universität Jena, seit Herbst 2002 Präsident des Zentrums für Humanwissenschaftliche Forschung. Arbeitsschwerpunkte: Theorie kausaler stochastischer Abhängigkeit, Latent-State-Trait-Modelle, Stochastische Messmodelle, Stimmungsmessung, Veränderungsmessung. Friedrich-Schiller-Universität Jena, Institut für Psychologie, Am Steiger 3/Haus 1, 07743 Jena. E-mail: [rolf.steyer@uni-jena.de](mailto:rolf.steyer@uni-jena.de); <http://www.uni-jena.de/svw/metheval/>

**Tabelle 1.** *Evaluation einer Behandlung.*

---

Erfolg	Behandlung		gesamt
	ja	nein	
ja	500	600	1100
nein	500	400	900
	1000	1000	2000

**Tabelle 2.** *Evaluation einer Behandlung getrennt nach Geschlecht.*

(a) Männer

Erfolg	Behandlung		gesamt
	ja	nein	
ja	300	75	375
nein	450	175	625
	750	250	1000

(b) Frauen

Erfolg	Behandlung		gesamt
	ja	nein	
ja	200	525	725
nein	50	225	275

**Tabelle 3.** Zahlenbeispiel, in dem für jede einzelne Person ein positiver Effekt der Experimentalgruppe besteht, und bei dem eine nicht ausgewogene Auswahl der Personen für die Experimental- bzw. Kontrollgruppe (s. die fett markierten Zahlen) zu verfälschten Schätzungen des durchschnittlichen Effekts führt.

(a)				(b)		
Person	$\mu_{E,u}$	$\mu_{K,u}$	$\mu_{E,u} - \mu_{K,u}$	$\mu_{E,u}$	$\mu_{K,u}$	$\mu_{E,u} - \mu_{K,u}$
$u_1$	82	68	14	<b>82</b>	68	14
$u_2$	89	81	8	<b>89</b>	81	8
$u_3$	101	89	12	<b>101</b>	89	12
$u_4$	108	102	6	<b>108</b>	102	6
$u_5$	118	112	6	118	<b>112</b>	6
$u_6$	131	119	12	131	<b>119</b>	12
$u_7$	139	131	8	139	<b>131</b>	8
$u_8$	152	138	14	152	<b>138</b>	14
Mittel	115	105	10	95	125	-30

(c)				(d)		
Person	$\mu_{E,u}$	$\mu_{K,u}$	$\mu_{E,u} - \mu_{K,u}$	$\mu_{E,u}$	$\mu_{K,u}$	$\mu_{E,u} - \mu_{K,u}$
$u_1$	<b>82</b>	68	14	82	<b>68</b>	14
$u_2$	89	<b>81</b>	8	<b>89</b>	81	8
$u_3$	<b>101</b>	89	12	<b>101</b>	89	12
$u_4$	<b>108</b>	102	6	108	<b>102</b>	6
$u_5$	118	<b>112</b>	6	<b>118</b>	112	6
$u_6$	<b>131</b>	119	12	131	<b>119</b>	12
$u_7$	139	<b>131</b>	8	139	<b>131</b>	8
$u_8$	152	<b>138</b>	14	<b>152</b>	138	14
Mittel	105.5	115.5	-10	115	105	10



**Tabelle 4.** Zahlenbeispiel, in dem für jede einzelne Person ein positiver Effekt der Experimentalgruppe gegenüber der Kontrollgruppe besteht und die Wahrscheinlichkeit, der Experimentalbedingung zugewiesen zu werden, unabhängig (vorletzte Spalte) bzw. abhängig (letzte Spalte) von der Person ist.

Person	$\mu_{E, u}$	$\mu_{K, u}$	$\mu_{E, u} - \mu_{K, u}$	$P(E u)$	$P(E u)$
$u_1$	82	68	14	1/2	8/9
$u_2$	89	81	8	1/2	7/9
$u_3$	101	89	12	1/2	6/9
$u_4$	108	102	6	1/2	5/9
$u_5$	118	112	6	1/2	4/9
$u_6$	131	119	12	1/2	3/9
$u_7$	139	131	8	1/2	2/9
$u_8$	152	138	14	1/2	1/9
			10		

*Anmerkung.* Die Wahrscheinlichkeit, eine der acht Personen zu ziehen, ist

**Tabelle 5.** Veranschaulichung einer homogenen Population.

Person	$\mu_{E, u}$	$\mu_{K, u}$	$\mu_{E, u} - \mu_{K, u}$
$u_1$	110	100	10
$u_2$	110	100	10
$u_3$	110	100	10
$u_4$	110	100	10
$u_5$	110	100	10
$u_6$	110	100	10
$u_7$	110	100	10
$u_8$	110	100	10
Mittel	110	100	10

**Tabelle 6.** *Beispiel für einen orthogonalen varianzanalytischen Versuchsplan.*

Therapie	Behandlungsbedürftigkeit			gesamt
	stark $Z = z_1$	mittel $Z = z_2$	schwach $Z = z_3$	
1 $X = x_1$	120 (10)	110 (20)	60 (10)	100 (40)
2 $X = x_2$	100 (10)	100 (20)	100 (10)	100 (40)
3 $X = x_3$	80 (10)	90 (20)	140 (10)	100 (40)
gesamt	100 (30)	100 (60)	100 (30)	100 (120)

**Tabelle 7.** Beispiel für einen nonorthogonalen varianzanalytischen Versuchsplan.

Therapie	Behandlungsbedürftigkeit			gesamt
	stark $Z = z_1$	mittel $Z = z_2$	Schwach $Z = z_3$	
1 $X = x_1$	120 (20)	110 (10)	60 (3)	(33)
2 $X = x_2$	100 (7)	100 (40)	100 (7)	(54)
3 $X = x_3$	80 (3)	90 (10)	140 (20)	(33)
Gesamt	(30)	(60)	(30)	(120)

Anmerkung. Zellenfrequenzen (in Klammern) und wahre Zellenmittelwerte.

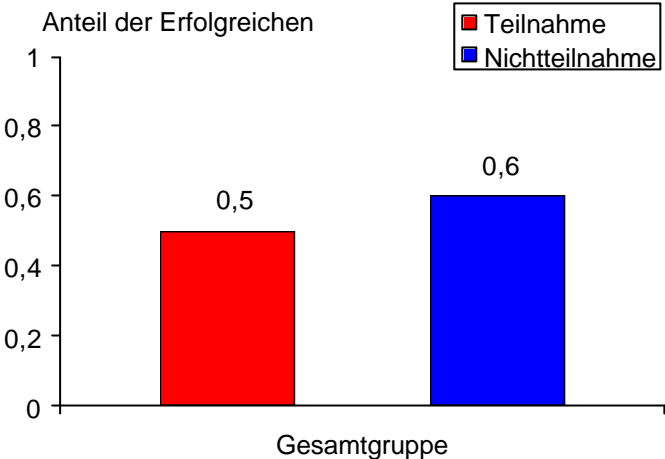
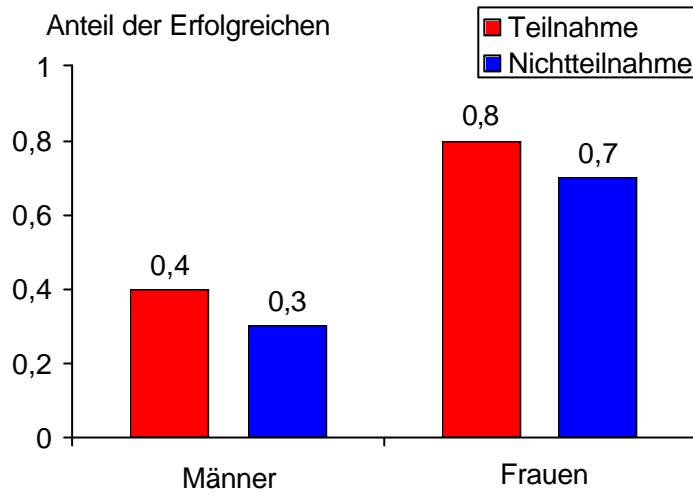


Abbildung 1. Histogramm der relativen Häufigkeiten für Erfolg in Tabelle 1.



**Abbildung 2.** Histogramm der relativen Häufigkeiten für Erfolg in Tabelle 2, getrennt nach Männern und Frauen.

Prima-facie-Effekt

$$PFE_{EK} := \mu_E - \mu_K$$

Individueller kausaler Effekt

$$ICE_{EK}(u) := \mu_{E,u} - \mu_{K,u}$$

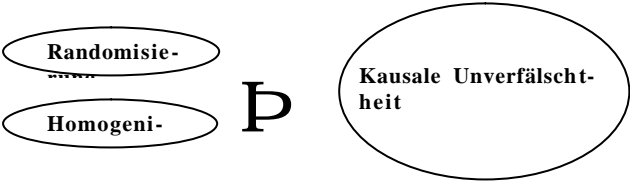
Durchschnittlicher kausaler Effekt

$$ACE_{EK} := \frac{1}{N} \sum_{i=1}^N (\mu_{E,u_i} - \mu_{K,u_i}).$$

Kausale Unverfälschtheit des Prima-facie-Effekts

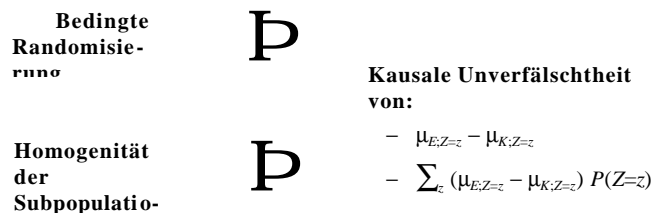
$$PFE_{EK} = ACE_{EK}.$$

**Abbildung 3.** Zusammenfassung der Grundbegriffe der Theorie individueller und durchschnittlicher kausaler Effekte.



**Abbildung 4.** *Hinreichende Bedingungen für die kausale Unverfälschtheit des Prima-facie-Effekts. Beide Bedingungen, Randomisierung und Homogenität, sind für sich alleine hinreichend für die kausale Unverfälschtheit.*





**Abbildung 5.** *Hinreichende Bedingungen für die kausale Unverfälschtheit der Prima-facie-Effekte in Subpopulationen und der adjustierten Effekte in der Gesamtpopulation.*



Abbildung 6. Implikationsstruktur der kausalen Unverfälschtheit des Prima-facie-Effekts in der Gesamtpopulation.