

Evaluation von Behandlungseffekten in quasi-experimentellen Studien

Evaluation of Treatment Effects in Quasi-Experimental Studies

Was wird erklärt

Je nachdem, welche statistischen Verfahren genutzt werden, können die Schätzungen von Behandlungseffekten in quasi-experimentellen Studien sehr unterschiedlich ausfallen (► Tab. 1; nähere Informationen zum Datenbeispiel: [1]). Die Verfahren unterscheiden sich darin, welche Einflussgrößen sie über die Behandlungsbedingungen hinaus auf welche Weise berücksichtigen. Auf welche Einflussgrößen muss hierbei geachtet werden? Und welche Verfahren werden der Komplexität der kausalen Abhängigkeiten in quasi-experimentellen Studien am ehesten gerecht?

Nur im Rahmen einer randomisierten Studie (zufällige Zuweisung von Probanden zu Behandlungen) kann erwartet werden, dass vor Beginn der Intervention jedes Merkmal in den Behandlungsbedingungen die gleichen Verteilungen aufweist (und damit auch gleiche Erwartungswerte, Varianzen etc.). Eine Erwartungswertdifferenz zwischen den Behandlungsbedingungen (z.B. behandelt versus nicht behandelt) bzgl. einer Outcomevariable kann hier eindeutig auf die Intervention zurückgeführt und daher als *durchschnittlicher Behandlungseffekt* interpretiert werden (d.h. als der zu erwartende Behandlungseffekt, wenn eine zufällig aus der Population ausgewählte Person behandelt versus nicht behandelt würde). In der medizinischen Psychologie und der Psychotherapieforschung ist eine randomisierte Zuweisung jedoch nicht immer möglich [2], etwa aus ethischen oder Praktikabilitätsgründen oder wenn Behandlungseffekte unter naturalistischen Bedingungen betrachtet werden sollen. Auch in solchen Fällen will man

die Frage untersuchen: Was ist der durchschnittliche Behandlungseffekt?

Quasi-experimentelle Studien

▼ In quasi-experimentellen Studien (keine zufällige Zuweisung zur Behandlung, stattdessen Selbstselektion der Probanden, z.B. nach Leidensdruck oder Fremdelektion, z.B. nach Symptomschwere durch einen zuweisenden Diagnostiker) muss damit gerechnet werden, dass sich Behandlungs- und Kontrollgruppe vor der Intervention in Merkmalen unterscheiden, von denen die Outcomevariable über die Intervention hinaus abhängt. Solche Merkmale werden konfundierende Variablen oder Störvariablen genannt. Mittelwertsunterschiede zwischen Behandlungsgruppen bzgl. der Outcomevariable können dann nicht als unverfälschte Schätzungen des Behandlungseffekts interpretiert werden. In quasi-experimentellen Studien ist es von entscheidender Bedeutung, mögliche konfundierende Merkmale zu erheben (im Weiteren bezeichnen wir ein gemessenes, potenziell konfundierendes Merkmal als Kovariate) und in der statistischen Auswertung adäquat zu berücksichtigen. Nachfolgend werden wichtige Aspekte der statistischen Analyse von Behandlungseffekten – insbesondere die generalisierte Kovarianzanalyse – erläutert. Zu den Fragen, welche potenziell konfundierenden Merkmale erhoben werden sollten und nach welcher Strategie Kovariaten ausgewählt werden können, s. [1, 3].

Häufig werden in quasi-experimentellen Studien Behandlungseffekte mit statistischen Verfahren untersucht, die für die Analyse randomisierter Experimente entwickelt wurden. Im Rahmen der 1-faktoriellen ANOVA etwa wird der durch-

schnittliche Behandlungseffekt geschätzt, indem die individuellen Outcomewerte in den Behandlungsgruppen gemittelt und zwischen den Behandlungsgruppen verglichen werden. Bei diesem Verfahren werden Kovariaten nicht berücksichtigt. Bei der 2-faktoriellen ANOVA mit einem Behandlungsgruppenfaktor und einem 2-stufigen Messwiederholungsfaktor werden die individuellen Prä-Post-Differenzen in den Behandlungsgruppen gemittelt und zwischen den Behandlungsgruppen verglichen. Wenn nicht sehr restriktive Annahmen gelten, führt dieses Verfahren zu verfälschten Schätzungen des durchschnittlichen Behandlungseffekts [4].

Traditionelle Kovarianzanalyse ANCOVA

▼ Bei der Berechnung des durchschnittlichen Behandlungseffekts in der traditionellen Kovarianzanalyse (ANCOVA) werden eine Kovariate Z oder ein Vektor $\mathbf{Z} = (Z_1, \dots, Z_m)$ von Kovariaten berücksichtigt, indem die ($Z=z$)-bedingten Mittelwertunterschiede in der Outcomevariable Y zwischen den Behandlungsgruppen geschätzt werden. Ein wesentlicher Vorteil dabei ist, dass bei der Schätzung des Behandlungseffekts die Kovariaten (z.B. Prätestwerte, Geschlecht etc.) statistisch kontrolliert werden. Die traditionelle ANCOVA beinhaltet allerdings auch eine bedeutsame Einschränkung. Es wird darin angenommen, dass keine Interaktion zwischen Kovariaten und Behandlungsgruppen besteht, d.h. dass die Behandlungseffekte für alle Kombinationen von Kovariatenwerten gleich sind.

Im einfachen Fall mit nur einer Kovariate Z , einer Outcomevariable Y und einer dichotomen Behandlungsvariable X (Kontrolle: $X=0$; Behandlung: $X=1$) wird in der traditionellen ANCOVA angenommen, dass sich die Regression von Y auf X und Z wie folgt darstellen lässt: $E(Y | X, Z) = \beta_0 + \beta_1 \cdot Z + \beta_2 \cdot X$. Hierbei ist β_2 die ($Z=z$)-bedingte Erwartungswertdifferenz in der Outcomevariable Y zwischen Kontroll- und Behandlungsbedingung, d.h. $\beta_2 = E(Y | X=1, Z=z) - E(Y | X=0, Z=z)$. Da in der traditionellen ANCOVA die ($Z=z$)-bedingten Erwartungswertdifferenzen für alle Werte von Z als gleich angenommen

Tab. 1 Behandlungseffekte am Datenbeispiel Bindungsängstlichkeit in Abhängigkeit vom verwendeten statistischen Verfahren (negative Effektgrößen kennzeichnen eine Verringerung der Bindungsängstlichkeit).

	Effektgröße	p-Wert
einfaktorielle ANOVA	0,324	<0,001
einfaktorielle ANOVA mit Messwiederholung	-0,450	<0,001
traditionelle ANCOVA	-0,084	0,317
g-ANCOVA Gesamtstichprobe $E[g_1(Z)]$	-0,076	0,221
g-ANCOVA Behandelte $E[g_1(Z) X=1]$	-0,200	0,015

werden (Abb. 1), ist β_2 auch die durchschnittliche Erwartungswertdifferenz und wird daher gewöhnlich als durchschnittlicher Behandlungseffekt interpretiert.

Dieser ist aber nur dann notwendigerweise mit dem oben definierten Behandlungseffekt identisch, wenn die Regression korrekt spezifiziert wurde (d. h. z. B. es besteht keine Interaktion zwischen Kovariate und Behandlungsgruppen) und die Regression unverfälscht ist. Letzteres ist z. B. dann gegeben, wenn keine weitere konfundierende Variable existiert. Ausführlich zu Kausalitätsbedingungen – d. h. Bedingungen, welche die Unverfälschtheit der Regression $E(Y|X, Z)$ implizieren – s. [5, 6].

Die generalisierte ANCOVA (g-ANCOVA)

Bei der g-ANCOVA werden Interaktionen zwischen Kovariaten und Behandlungsgruppen zugelassen, d. h. in Abhängigkeit von den Kovariatenwerten können Behandlungseffekte größer oder kleiner sein. Im einfachen Fall mit nur einer Kovariate Z und einer dichotomen Behandlungsvariable X (Kontrolle: $X=0$; Behandlung: $X=1$) kann man die Regression von Y auf X und Z immer wie folgt darstellen: $E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X$, wobei man g_0 als *Achsenabschnittsfunktion* und g_1 als *Effektfunktion* bezeichnet. Sind wir an den Effekten von X auf Y interessiert, dann gibt ein einzelner Wert $g_1(z)$ den Effekt von X auf Y bei gegebenem Wert z der Kovariate Z an. Diese ($Z=z$)-bedingten Behandlungseffekte können für jeden Wert z von Z verschieden sein. So zeigt sich in unserem Beispiel (Abb. 2), dass die Behandlungseffekte (s. Abstände zwischen den beiden Regressionsgeraden) mit größer werdendem Z stärker in die erwünschte Richtung weisen (geringere Post-Bindungsängstlichkeit der Patienten im Vergleich zu Kontrollprobanden) und sich unterhalb eines Wertes von $z=1,1$ zunehmend in die unerwünschte Richtung entwickeln (höhere Post-Bindungsängstlichkeit der Patienten im Vergleich zu Kontrollprobanden). Gilt das Interesse nicht nur den ($Z=z$)-bedingten, sondern auch dem durchschnittlichen Behandlungseffekt, so betrachtet man den Erwartungswert $E[g_1(Z)]$ der Effektfunktion $g_1(Z)$ über die Verteilung von Z .

Die Achsenabschnitts- und die Effektfunktionen $g_0(Z)$ und $g_1(Z)$ können auf verschiedene Weisen parametrisiert werden.

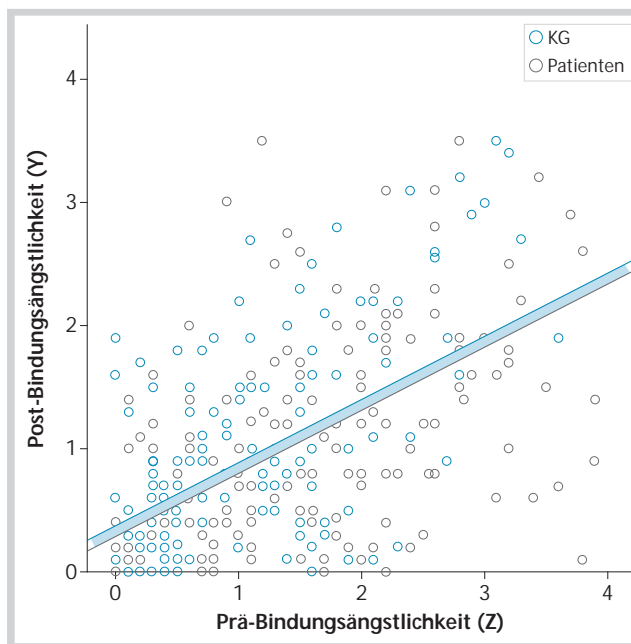


Abb. 1 Grafische Veranschaulichung des anhand der traditionellen ANCOVA berechneten Behandlungseffekts (Abstand zwischen den beiden Regressionsgeraden) am Datenbeispiel Bindungsängstlichkeit. Die graue Linie entspricht der Regressionsgeraden der Post- auf die Prä-Bindungsängstlichkeit in der Gruppe der Patienten, die blaue der in der Kontrollgruppe.

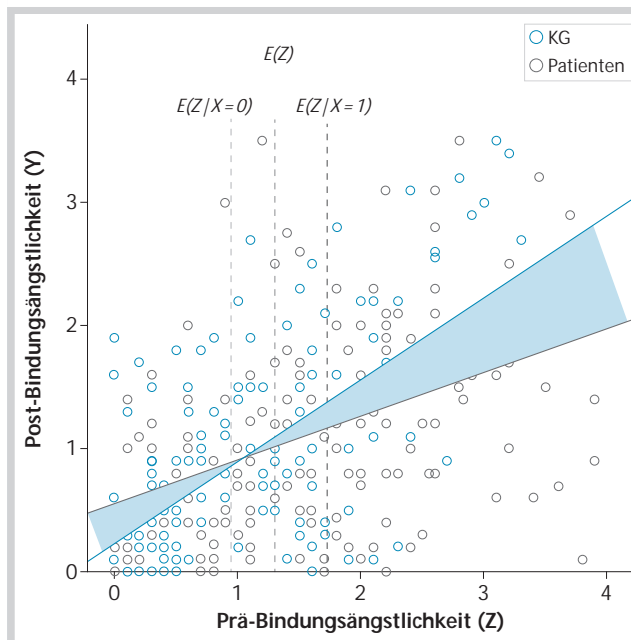


Abb. 2 Grafische Veranschaulichung der anhand der g-ANCOVA berechneten bedingten Behandlungseffekte (Abstände zwischen den beiden Regressionsgeraden) am Datenbeispiel Bindungsängstlichkeit. Die graue Linie entspricht der Regressionsgeraden der Post- auf die Prä-Bindungsängstlichkeit in der Gruppe der Patienten, die blaue der in der Kontrollgruppe.

Liegt eine einzige numerische Kovariate vor, so ist die einfachste Parametrisierung $g_0(Z) = \gamma_{00} + \gamma_{01} \cdot Z$ und $g_1(Z) = \gamma_{10} + \gamma_{11} \cdot Z$.¹ In diesem Fall lässt sich die Regression von Y auf X und Z wie folgt schreiben:

$$E(Y|X, Z) = g_0(Z) + g_1(Z) \cdot X \quad (1)$$

$$= (\gamma_{00} + \gamma_{01} \cdot Z) + (\gamma_{10} + \gamma_{11} \cdot Z) \cdot X \quad (2)$$

Für den durchschnittlichen Behandlungseffekt gilt dann:

$$E[g_1(Z)] = E(\gamma_{10} + \gamma_{11} \cdot Z) = \gamma_{10} + \gamma_{11} E(Z) \quad (3)$$

¹ In komplexeren Modellen können g_0 und g_1 auch als Polynome höherer Ordnung dargestellt werden. Komplexere Fälle mit mehreren Kovariaten $Z = (Z_1, \dots, Z_m)$ und mehreren Behandlungsgruppen (x_0, x_1, \dots, x_j) kann man durch Erweiterung der Gleichung 1 bzw. 2 parametrisieren (für Details s. [7]).

Da der Erwartungswert $E[g_1(Z)]$ der Durchschnitt der ($Z=z$)-bedingten Behandlungseffekte ist, ist er auch der Durchschnitt der personenspezifischen Behandlungseffekte und kann als der zu erwartende Behandlungseffekt bei einer zufällig aus der Gesamtpopulation gezogenen Person interpretiert werden. Dieser Behandlungseffekt ist unverfälscht, wenn die Regression $E(Y|X, Z)$ korrekt spezifiziert und unverfälscht ist.

Die g-ANCOVA mittels EffectLite

EffectLite [7] ist eine kostenfrei zugängliche Software, die als Präprozessor Inputfiles für LISREL und Mplus erstellt, die in das entsprechende Programm eingelesen werden und der Berechnung einer g-ANCOVA dienen. Um EffectLite-Analysen durchführen zu können, muss entsprechend LISREL oder Mplus auf dem Computer installiert sein.

Vorteile

Die g-ANCOVA mittels EffectLite weist nicht nur den Vorteil auf, dass Kovariaten-Behandlungsgruppen-Interaktionen modelliert werden. Weitere Vorteile sind u.a.:

- (1) Behandlungseffekte werden nicht nur für die Gesamtpopulation geschätzt, sondern es werden auch $(X=x)$ -bedingte Behandlungseffekte gegeben die einzelnen Behandlungsgruppen x berechnet, also $E[g_1(Z)|X=x]$ (Näheres hierzu im nächsten Abschnitt).
- (2) Kovariaten sowie Outcome können als latente (messfehlerbereinigte) Variablen spezifiziert werden.
- (3) Varianzhomogenität zwischen den Gruppen ist keine notwendige Bedingung für die Validität der Inferenzstatistiken.
- (4) Es werden nicht nur durchschnittliche Behandlungseffekte angegeben, sondern auf Grundlage des individuellen Musters von Kovariatenwerten auch Behandlungseffekte für jeden einzelnen Probanden.

$(X=x)$ -bedingte Behandlungseffekte

Die Berechnung $(X=x)$ -bedingter Behandlungseffekte ist bei der Analyse von quasi-experimentellen Interventionsstudien von besonderer Bedeutung. Betrachten wir wieder den einfachen Fall mit nur einer Kovariate Z und einer dichotomen Behandlungsvariable X (Kontrollgruppe: $X=0$; Behandlungsgruppe: $X=1$). Wie bereits erwähnt, wird in der g-ANCOVA der durchschnittliche Behandlungseffekt als Erwartungswert der Funktion $g_1(Z)$ über die Verteilung von Z ermittelt (vgl. Gleichung (3)). Gehen wir wieder von einer linearen Effektfunktion $g_1(Z)$ von Z aus und besteht eine Kovariaten-Behandlungsgruppen-Interaktion (d.h. gilt $\gamma_{11} \neq 0$), so ist der durchschnittliche Behandlungseffekt abhängig vom Erwartungswert $E(Z)$. Im randomisierten Experiment kann davon ausgegangen werden, dass der (unbedingte) Erwartungswert $E(Z)$ – d.h. der Erwartungswert von Z über die Gesamtstichprobe – gleich den $(X=x)$ -bedingten Erwartungswerten von Z – d.h. den

behandlungsgruppenspezifischen Erwartungswerten $E(Z|X=1)$ und $E(Z|X=0)$ – ist, da Randomisierung impliziert, dass die Erwartungswerte der Kovariaten zwischen den Behandlungsgruppen gleich sind. Im Quasi-Experiment dagegen sind die Kovariaten zwischen den Gruppen häufig sehr unterschiedlich verteilt (Abb. 2), sodass gilt: $E(Z) \neq E(Z|X=1) \neq E(Z|X=0)$. Folglich kann sich der (unbedingte) durchschnittliche Behandlungseffekt $E[g_1(Z)]$ deutlich unterscheiden von den $(X=x)$ -bedingten durchschnittlichen Behandlungseffekten, die in EffectLite mitberechnet werden. Der $(X=x)$ -bedingte durchschnittliche Behandlungseffekt für die Behandelten $E[g_1(Z)|X=1] = \gamma_{10} + \gamma_{11} E(Z|X=1)$ ist eine Schätzung für den zu erwartenden Behandlungseffekt, wenn eine Person zufällig aus der Subpopulation der Behandelten (im Beispiel: der Patienten) ausgewählt wird. Analog ist $E[g_1(Z)|X=0] = \gamma_{10} + \gamma_{11} E(Z|X=0)$ der zu erwartende Behandlungseffekt für eine Person, die zufällig aus der Kontrollgruppenpopulation gezogen wird.

Wahlmöglichkeit

Bei der Verwendung von EffectLite kann der Forscher entscheiden, ob entsprechend seiner Hypothesen einer der $(X=x)$ -bedingten Behandlungseffekte oder der unbedingte Behandlungseffekt $E[g_1(Z)]$ von Interesse ist. Bei den traditionellen statistischen Verfahren, wie ANOVA oder ANCOVA, besteht diese Wahlmöglichkeit nicht. In quasi-experimentellen Interventionsstudien ist oftmals – wenn nicht in der Regel – der $(X=1)$ -bedingte Behandlungseffekt, d.h. der Behandlungseffekt bedingt auf die Gruppe der Behandelten $E[g_1(Z)|X=1]$, der tatsächlich interessierende, da in der Kontrollgruppe und damit implizit in der Gesamtstichprobe viele Probanden enthalten sind, die entsprechend des zugrunde liegenden Selektionsprozesses nur eine geringe Behandlungswahrscheinlichkeit haben – weil z.B. nur eine geringe Behandlungsbedürftigkeit (im Beispiel eine geringe Bindungsängstlichkeit) vorliegt.

Literatur

- 1 Kirchmann H, Steyer R, Mayer A et al. Effects of adult inpatient group psychotherapy on attachment characteristics: An observational study comparing routine care to an untreated comparison group. *Psychother Res* 2012; 22: 95–114
- 2 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312: 1215–1218
- 3 Steiner PM, Cook TD, Shadish WR et al. The importance of covariate selection in control-

Fazit für die Praxis

Liegen zwischen Behandlungs- und Kontrollgruppe unterschiedliche Kovariatenverteilungen vor und bestehen zudem Kovariaten-Behandlungsgruppen-Interaktionen, wird durch ANOVA und traditionelle ANCOVA der Behandlungseffekt nicht nur falsch geschätzt, es besteht sogar die Gefahr, unwissentlich eine falsche Fragestellung zu analysieren. Für die Analyse quasi-experimenteller Interventionsstudien empfehlen wir die Verwendung der generalisierten Kovarianzanalyse mittels EffectLite und gegebenenfalls die Betrachtung des $(X=1)$ -bedingten durchschnittlichen Behandlungseffekts – d.h. des durchschnittlichen Behandlungseffekts, der auf der Basis der Kovariatenverteilungen in der Behandlungsgruppe ermittelt wird.



Dr. Helmut Kirchmann

Institut für Psychosoziale Medizin und Psychotherapie, Universitätsklinikum Jena, Friedrich-Schiller-Universität Jena, Stoystraße 3, 07740 Jena, E-Mail: helmut.kirchmann@med.uni-jena.de



Prof. Rolf Steyer

Lehrstuhl für Methodenlehre und Evaluationsforschung, Institut für Psychologie, Friedrich-Schiller-Universität Jena, Am Steiger 3, Haus 1, 07743 Jena, E-Mail: rolf.steyer@uni-jena.de

ling for selection bias in observational studies. *Psychol Methods* 2010; 15: 250–267

- 4 Breukelen GJP. ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. *J Clin Epidemiol* 2006; 59: 920–925
- 5 Steyer R, Gabler S, von Davier A et al. Causal regression models II: Unconfoundedness and causal unbiasedness. *Methods of Psychological Research Online* 2000; 5: 55–86
- 6 Steyer R, Nachtigall C, Wüthrich-Matrone O et al. Causal regression models III: Covariates, conditional, and unconditional average causal effects. *Methods of Psychological Research Online* 2002; 7: 41–68
- 7 Steyer R, Partchev I. EffectLite for LISREL: A program for the uni- and multivariate analysis of unconditional, conditional and average mean differences between groups [computer software]. Universität Jena. Im Internet: www.statlite.com. Stand: 25.6.2008