

# Experimentelle Methoden

von Stefan Klonner

## 1) Allgemeine Informationen

Experimentelle Methoden gehören zur Gruppe der ökonometrischen oder quantitativen Verfahren der Wirkungsanalyse. Im Vordergrund steht die Bestimmung eines Wirkungseffekts (*treatment effect*) einer bestimmten Massnahme. Von primärem Interesse ist hierbei der *individuelle Wirkungseffekt*, welcher definiert ist als der Unterschied zwischen zwei Szenarien, zum einen der Situation eines Individuums bei Teilnahme an einer bestimmten Maßnahme, zum anderen der Situation des selben Individuums, wenn die Massnahme nicht durchgeführt worden wäre. Der *durchschnittliche Wirkungseffekt* ist hingegen der Durchschnitt der individuellen Wirkungseffekte in einer gegebenen Population von Individuen. Die grundlegende Eigenschaft von experimentellen Verfahren ist, dass Teilnehmer (*treatment group*) und Nicht-Teilnehmer (*control group*) an einer Massnahme zufällig, vor Durchführung der Maßnahme, bestimmt werden und die *output*- und *outcome*-Variablen von Interesse in beiden Gruppen vor und nach Durchführung der Maßnahme erhoben werden. Für jede solche Variable läßt sich damit der durchschnittliche Wirkungseffekt ermitteln, der den entsprechenden Output bzw. Outcome der Massnahme quantifiziert. Die methodische Bedeutung experimenteller Verfahren liegt darin begründet, daß nur sie eine gültige (genauer *konsistente*) Schätzung durchschnittlicher Wirkungseffekte erlauben. Wenn bei der Projektplanung kein experimenteller Ansatz zum Einsatz kommt, unterscheiden sich *treatment*- und *control*-Gruppe im Allgemeinen in einer Weise, die eine gültige Schätzung eines Wirkungseffektes unmöglich macht. Entsprechend kommen alle anderen quantitativen, nicht-experimentellen Verfahren, die der Schätzung von Wirkungseffekten dienen, nicht ohne zusätzliche Annahmen aus (z. B. über die Vergleichbarkeit von *treatment* und *control* Individuen; siehe hierzu den Artikel über mikroökonomische Verfahren).

Wegen ihrer Objektivität aus methodologischer Sicht sind experimentelle Methoden im Verlauf der letzten Jahre zu einer Art Goldstandard der Evaluierung entwicklungspolitischer Massnahmen geworden. Bei Projekten der Weltbank, verschiedenen nationalen Programmen und zahlreichen Nicht-Regierungs-Organisationen kommen sie regelmäßig zur Anwendung. Darüber hinaus erfreuen sich experimentelle Verfahren bei akademischen Ökonomen größter Beliebtheit. So sind Entwicklungsökonomien führender US-amerikanischer Universitäten (Harvard, MIT, Yale, Berkeley) an der Evaluation zahlloser Weltbank- und NGO-Projekte beteiligt.

## 2) Anwendungsbereiche

Eine grundlegende Voraussetzung der Anwendung experimenteller Verfahren ist, dass zufällig ausgewählte *treatment*- und *control*-Individuen in ausreichender Anzahl zur Verfügung stehen und dass die Kontrollgruppe keinen, auch keinen indirekten, Effekten der Maßnahme unterliegt. Diese Anforderungen machen sofort klar, dass sich experimentelle Methoden nicht zur Evaluierung von Massnahmen eignen, die auf der Makroebene angesiedelt sind, z. B. Stärkung der Zentralbankunabhängigkeit. Hingegen sind Projekte auf der Mikroebene und z. T. auf der Mesoebene gut geeignet. In der Vergangenheit wurden Gesundheits- und Erziehungsmassnahmen (sh. z. B. die Studie über Schulbücher in Kenia von Glewwe, Kremer und Moulin, 2004), Programme zur Förderung kleinster und kleiner Unternehmer, Entwicklung des ländlichen Raums, aber auch Reformen der Kommunalregierung erfolgreich evaluiert.

## 3) Analyseansatz

Im Vordergrund der experimentellen Verfahren steht die quantitative Bestimmung durchschnittlicher Wirkungseffekte. Bei entsprechend sorgfältiger Konzeption und Anwendung des Evaluationsverfahrens lassen sich darüber hinaus auch hypothetische Wirkungsketten überprüfen. Der hauptsächliche Evaluations-Output besteht aus Tabellen, in denen für verschiedene Variablen die der Massnahme zuschreibenden Veränderungen sowie deren Standardfehler aufgeführt sind. Dies erlaubt neben der Schätzung von durchschnittlichen Effekten eine Beantwortung der Frage, ob eine Massnahme einen statistisch signifikanten Effekt auf eine bestimmte Größe hat. Diese einfache, quantitative Darstellung der Ergebnisse erlaubt einen einfachen Vergleich mit der Wirkung verwandter oder alternativer Massnahmen und kann somit eine nützliche Grundlage für größer angelegte Kosten-/Nutzenanalysen bilden.

## 4) Durchführung (inkl. Datenerhebung und Datenauswertung)

### a) Verfahrensablauf

Experimentelle Methoden bei der Evaluierung setzen eine experimentelle Anlage der Massnahme voraus. Hiermit ist klar, dass die Massnahme und ihre Evaluation simultan geplant und durchgeführt

werden müssen. Wichtigster Punkt hierbei ist die zufällige Auswahl der Teilnehmer und der Kontrollindividuen vor Beginn der Massnahme. Es folgt ein schematischer Ablauf einer auf experimentellen Methoden basierenden Evaluation.

1. *Projekt- und Evaluationsplanung:* Vor Beginn der Massnahme findet zweierlei statt. Zum einen werden Variablen identifiziert, auf die das geplante Projekt möglicherweise einen Einfluss hat. Die Auswahl der Variablen, deren Änderung gemessen werden soll, kann aus einer vorhergehenden Erstellung hypothetischer Wirkungsketten resultieren. Solche Überlegungen sind insbesondere zur Vermeidung von späteren Zuordnungslücken erforderlich. Die Variablen können quantitativer oder qualitativer Art sein. Zum anderen werden die *treatment*- und die *control*-Gruppe ausgewählt. Diese werden aus einer Grundgesamtheit von Beobachtungseinheiten zufällig („randomisiert“) bestimmt. Wichtig ist, die Ebene der Beobachtungseinheiten, auf der randomisiert wird, so zu wählen, dass *control*-Individuen nicht von der Massnahme beeinflusst werden. Z. B. ist bei Bestehen von Externalitäten oder Spillover Effekten zwischen Individuen wie bei einem Schulbücherprogramm eine Randomisierung auf Schul-Ebene (d.h. 50% der Schulen erhalten Bücher, 50% keine) sinnvoller als auf individueller Ebene innerhalb von Schulen (in jeder Schule erhalten genau 50% der Schüler Bücher), da im zweiten Szenario die Kontroll-Individuen von den Büchern ihrer Mitschüler profitieren könnten. Dieser Spillover-Effekt liesse sich aber nicht messen und eine Unterschätzung des durchschnittlichen Wirkungseffekts könnte die Folge sein. Wenn die statistische Signifikanz der Ergebnisse von Bedeutung ist (z. B.: „Eine Verbesserung der Lesefähigkeit von 5% oder mehr soll mit 95-prozentiger Wahrscheinlichkeit als statistisch signifikant identifiziert werden), kann die erforderliche Stichprobengröße mit Hilfe von Güteberechnungen ermittelt werden. Hierzu sind Daten über die Verteilung (Mittelwert und Varianz) der betrachteten Variablen in der Population erforderlich (sh. Duflo et al., 2008, Abschnitt 4). Ist eine Disaggregation der Ergebnisse (z. B. nach Gender) geplant, muß dies auch bei der Stichprobengröße berücksichtigt werden.
2. *Baseline Survey:* Vor Durchführung der Massnahme werden die Variablen von Interesse bei allen Individuen der Stichprobe erhoben. Die Randomisierung stellt im Prinzip sicher, dass sich die *treatment*- und *control*-Individuen nicht in statistisch signifikanter Weise unterscheiden. Dies sollte aufgrund der Baseline Daten überprüft werden.
3. *Durchführung der Massnahme,* eventuell begleitende Datenerhebungen in der Stichprobe, gemäß Schritt 1 (z. B. Erhebung der Nutzung der ausgeteilten Schulbücher bei den *treatment*-Schülern).
4. *Follow-up Survey:* Nach Durchführung der Massnahme erneute Erhebung der Variablen von Interesse bei allen Individuen der Stichprobe.
5. *Statistische Auswertung:* Der durchschnittliche Wirkungseffekt auf Variable Y (z. B. Punktzahl bei einem Lesetest) wird berechnet als doppelte Differenz,  $(\bar{y}_1^{treat} - \bar{y}_0^{treat}) - (\bar{y}_1^{control} - \bar{y}_0^{control})$ , wobei  $\bar{y}_t^{group}$  den Stichprobenmittelwert der Untergruppe group = {treat, control} in Survey-Runde t (0 für Baseline, 1 für Follow-up) bezeichnet. Mit Hilfe elementarer Methoden der linearen Regressionsanalyse lassen sich darüber hinaus auf einfache Weise die statistischen Eigenschaften des durchschnittlichen Wirkungseffekts bestimmen.

## **b, c, d) Akteure, Datenerhebung und -auswertung**

Die Datenerhebung kann vom Projektteam oder einer lokalen Survey-Organisation durchgeführt werden. Um die Objektivität der Ergebnisse zu gewährleisten, ist eine Trennung der Projekt-Durchführung einerseits und der Datenerhebung und -auswertung andererseits jedoch wünschenswert.

### **5) Nutzung der Ergebnisse**

Experimentelle Methoden betonen die Konsistenz und Objektivität der gewonnenen statistischen Ergebnisse. Entsprechend bieten die Ergebnisse nicht nur der durchführenden Organisation oder dem Geber einen Überblick über die Wirkung der Massnahme. Insbesondere kann das Wissen um die Wirkung bzw. den Erfolg bestimmter Massnahmen als öffentliches Gut für die Gestaltung zukünftiger Massnahmen angesehen werden. Entsprechend argumentieren Befürworter dieses Ansatzes, dass auf experimentellen Methoden beruhende Evaluationsergebnisse in einer Datenbank gesammelt und als Grundlage künftiger Projektplanung dienen sollten. Ansätze hierzu sind bei der Weltbank vorhanden. Die Organisation *International Initiative for Impact Evaluation* (3ie) verfolgt explizit dieses Ziel und bietet eine Plattform, auf der entsprechende Ergebnisse online verfügbar gemacht werden.

### **6) Anforderungen**

Die Anwendung experimenteller Methoden ist im Allgemeinen aufwändig. Sie erfordert eine randomisierte Durchführung der Massnahme und eine genaue Vorab-Planung. Eine Randomisierung der Massnahme stößt oft auf praktikable oder politische Widerstände. Datenerhebungen müssen zu zwei, üblicherweise weit auseinanderliegenden Zeitpunkten durchgeführt werden. Design und Auswertung der Daten erfordern erhebliche statistische Expertise. Der genaue zeitliche Aufwand im Feld richtet sich nach Umfang des Fragebogens und der Größe der Stichprobe.

### **7) Besondere Eignung und Grenzen des Verfahrens**

Experimentelle Verfahren eignen sich besonders zur Evaluierung von Projekten auf der Mikroebene, wenn es um quantitativ messbare Effekten geht. Der Fokus liegt auf der Wirkungsebene (outcome und impact). Die Stärke des Ansatzes liegt in der methodischen Stichhaltigkeit. Die Ermittlung von Wirkungsketten und Zuordnung bestimmter Effekte sind dagegen nur in dem Ausmaß möglich, wie dies bei der Vorabplanung berücksichtigt wurde.

Es bestehen darüber hinaus weitere Kritikpunkte an dieser Methode. Zwar werden die Effekte für das Umfeld, in dem die Massnahme stattfindet, objektiv gemessen. Jedoch läßt dies nicht unbedingt eine Verallgemeinerung auf andere Umfelder zu. Weiterhin können beim *scaling-up* einer Massnahme, die im Pilotstadium mit experimentellen Verfahren evaluiert wurde, zusätzliche Gleichgewichtseffekte eine Rolle auftreten, die in der Pilotstudie keine Rolle spielten. Schliesslich verhalten sich Individuen möglicherweise anders, wenn ihnen bewußt ist, dass sie Teil eines Experiments sind, als wenn die Massnahme für alle oder keinen durchgeführt würde.

### **Literaturhinweise**

- Duflo, Esther, Rachel Glennerster and Michael Kremer, 2008. Using Randomization in Development Economics Research: A Toolkit. Chapter 61 in T. Paul Schultz and John Strauss (eds.) *Handbook of Development Economics*, vol. 4, pp 3895-3962. Elsevier: North-Holland.
- Duflo, Esther and Michael Kremer, 2005. Use of Randomization in the Evaluation of Development Effectiveness. In George Pitman, Osvaldo Feinstein and Gregory Ingram (eds.) *Evaluating Development Effectiveness*. New Brunswick, NJ: Transaction Publishers 2005, pp. 205-232.
- Glewwe, Paul, Michael Kremer, and Sylvie Moulin, 2004). Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya, Mimeo, Harvard University.
- Kremer, Michael, 2003. Randomized Evaluations of Educational Programs in Developing Countries: Some Lessons, *American Economic Review Papers and Proceedings* 93(2): 102-115.
- Ravallion, Martin, 2001. The Mystery of the Vanishing Benefits: An Introduction to Impact Evaluation, *World Bank Economic Review* 15(1), 115-14.
- Ravallion, Martin 2008). Evaluating Anti-Poverty Programs. Chapter 59 in T. Paul Schultz and John Strauss (eds.) *Handbook of Development Economics*, vol. 4, pp 3787-3846. Elsevier: North-Holland.