

Empirische Revolution in der Verwaltung

Datenbasierte Wirkungsanalysen halten weltweit Einzug in der öffentlichen Verwaltung. In der Schweiz gehört unter anderem die Direktion für Entwicklung und Zusammenarbeit (Deza) zu den Vorreiterinnen. *Dina Pomeranz*

Abstract Wirkungsanalysen greifen seit einigen Jahren auf präzisere Methoden und umfangreiche Datenquellen zurück. Dies stellt eine grosse Chance für öffentliche Verwaltungen, internationale Organisationen und Firmen dar. Weltweit verändert sich die Art und Weise, wie sich Institutionen Informationen über die Wirkung ihres Vorgehens erarbeiten. Indem Verwaltungen eng mit Forschenden aus den Wirtschafts- und Sozialwissenschaften zusammenarbeiten, entstehen methodische Innovationen. Die evaluierten Themenbereiche reichen von Bildungs- und Gesundheitspolitik über Beschaffungswesen und Steuerverwaltung bis zu Integrations- und Arbeitsmarktpolitik.

Mit der wachsenden Verfügbarkeit grosser Datenquellen («Big Data») werden Wirkungsanalysen zudem immer kostengünstiger. Die Verwaltung besitzt eine breite Palette an administrativen Daten, welche dazu verwendet werden können, Wirkungen von Reformen und Interventionen zu messen. Dadurch sind keine teuren Umfragen nötig, was die Kosten erheblich senkt.

Die täglichen Entscheidungen der öffentlichen Verwaltung beeinflussen die wirtschaftliche und gesellschaftliche Entwicklung eines Landes: Welches Mittel wirkt am besten gegen Steuerhinterziehung? Welche Massnahmen führen dazu, dass Vorschriften besser eingehalten werden? Wie reagieren Firmen auf Steueranreize, und wann funktioniert die Unfallprävention am besten?

Die Wirkung solcher Massnahmen zu messen, ist in der Realität allerdings nicht ganz einfach. In den letzten Jahren hat eine stille Revolution stattgefunden in der Verbesserung und Anwendung von Methoden zur Wirkungsanalyse. Neue Methoden erlauben es, kausale Zusammenhänge immer besser zu messen und zu verstehen. In zunehmendem Masse arbeiten die öffentlichen Verwaltungen mit Forschenden aus der Wissenschaft bei der Erarbeitung von entsprechenden Studien zusammen. Dies hat auch zur rapiden Verbreitung dieser Methoden beigetragen. Die Kombination von praktischer Expertise aus der Verwaltung mit methodischer Expertise aus der Volkswirtschaftsforschung und aus anderen Sozialwissenschaften ist dabei besonders wertvoll.

Wie misst man den «Kontrafakt»?

Bei jeder statistischen Analyse ist es zentral, dass die gewählte Methode an die zu analysierende Frage angepasst ist. Soll die Wirkung einer Reform oder eines Programms auf bestimmte Ergebnisse (Outcomes) untersucht werden, dann sind Methoden gefragt, die kausale Effekte verlässlich messen. Wie wirkt sich beispielsweise ein Mahnbrief auf die Steuerzahlungen aus? Oder was ist der Effekt von Lohnerhöhungen für Lehrpersonen auf die Anzahl neuer Lehrkräfte?

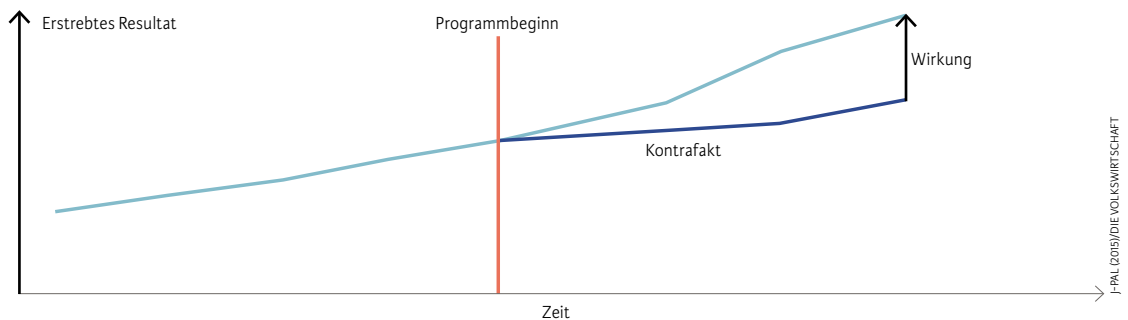
Um solche kausalen Wirkungsanalysen durchzuführen, stehen mehrere Methoden zur Verfügung. Allen gemeinsam ist, dass sie untersuchen, was eine Reform, ein Programm, eine Intervention oder eine Politik ausgelöst haben. Dabei besteht die grundlegende Herausforderung jeweils darin, dass in der Realität nicht beobachtet werden kann, was ohne die Reform geschehen wäre. Diese hypothetische Situation dessen, was ohne die Intervention geschehen wäre, wird als «Kontrafakt» bezeichnet (siehe *Abbildung auf Seite 6*).

Da der Kontrafakt in der Realität nicht direkt beobachtet werden kann, wird bei jeder Wirkungsanalyse explizit oder implizit versucht, diesen so gut wie möglich einzuschätzen.



Die Verwaltung besitzt unzählige Daten. Digitale Tools vereinfachen die Wirkungsanalysen.

Was wäre ohne Intervention passiert?



Die Grafik stellt die grundlegende Herausforderung der Wirkungsmessung dar, bei der versucht wird, den Unterschied zwischen dem tatsächlich eingetretenen Ergebnis (hellblau) und dem Kontrafakt (dunkelblau) zu messen.

Je präziser die Einschätzung, umso besser die Qualität der Wirkungsmessung. Normalerweise wird dazu eine Kontroll- oder Vergleichsgruppe beigezogen.¹

Eine Kontrollgruppe besteht aus Personen oder Firmen, die nicht vom Programm oder der Reform betroffen waren. Um die Wirkung zu messen, werden die Ergebnisse der betroffenen Gruppe mit den Ergebnissen der Kontrollgruppe verglichen. Je stärker die Eigenschaften der Personen der Kontrollgruppe mit denjenigen der Betroffenen übereinstimmen, desto verlässlicher sind die Ergebnisse der Wirkungsstudie. Sprich: Wenn beide Gruppen ohne Intervention gleich wären und sich gleich verhielten, könnte der Unterschied in den Ergebnissen auf das zu evaluierende Programm oder die Reform zurückgeführt werden.

Selbstselektion verzerrt Resultate

Warum ist die Konstruktion einer guten Vergleichsgruppe so schwierig? Oft unterscheiden sich die Teilnehmenden eines Programms oder die Betroffenen einer Politik auf eine Art und Weise vom Rest der Bevölkerung, die einen Vergleich schwierig macht. Nehmen wir zum Beispiel eine Massnahme im Arbeitslosenbereich, die darauf abzielt, die Chancen bei der Stellensuche zu erhöhen. In diesem Fall könnte man die Arbeitslosengruppen in zwei Gruppen einteilen: Die erste Gruppe hat sich entschieden, bei der Massnahme mitzumachen, die zweite nicht.

Wenn sich nun herausstellt, dass Erstere schneller eine Stelle finden, ist unklar, ob dies

aufgrund der Massnahme oder aus anderen Gründen geschah. Waren die Teilnehmenden vielleicht motivierter und haben deshalb teilgenommen? Möglicherweise fanden sie aufgrund der höheren Motivation – und nicht wegen der Massnahme – eine Stelle. Eine Selbstselektion kann somit die Messung verzerren und zu falschen Schlussfolgerungen führen.

Die wohl berühmteste Methode, die zum Ziel hat, solche Selektionseffekte zu vermeiden, ist die randomisierte Feldstudie: Ähnlich wie bei medizinischen Studien für neue Medikamente werden Individuen, Firmen, Schulklassen oder andere Gruppen nach dem Zufallsprinzip verschiedenen Bedingungen ausgesetzt. Wenn die Gruppe der Teilnehmenden gross genug ist, lässt sich die Wirkung der verschiedenen Bedingungen verlässlich messen. So kann man beispielsweise analysieren, wie sich unterschiedliche Mahnbriefe der Steuerbehörde auf die Steuermoral auswirken oder wie verschiedene Integrationsmassnahmen die gesellschaftliche und wirtschaftliche Teilnahme von Zugewanderten beeinflussen. Damit randomisierte Feldstudien funktionieren, müssen sie sorgfältig im Voraus geplant werden.

Eine weitere, ebenfalls häufig angewandte Methode ist die Regressions-Diskontinuitäts-Analyse (Regression Discontinuity Design, RDD). Dieses Verfahren bietet sich an, wenn die Teilnahme an einem Programm oder die Betroffenheit von einer Reform von einer klaren numerischen Eintrittsgrenze abhängt. Ein Beispiel: Eine Steuerbehörde wählt nach dem Abgabetermin der Steuererklärung Steuerzahlende

¹ Vgl. Pomeranz (2017).

mit Abzügen von über 15 Prozent für eine zusätzliche Steuerprüfung aus. Steuerzahlende mit Abzügen von 15,1 Prozent sind also dieser zusätzlichen Überprüfung ausgesetzt, diejenigen mit 14,9 Prozent nicht.

Die Regressions-Diskontinuitäts-Analyse erlaubt es, die Wirkung der Prüfung zu analysieren, indem Steuerzahlende knapp über der 15-Prozent-Schwelle mit denjenigen knapp darunter verglichen werden. Ähnlich könnte man zum Beispiel die Wirkung eines bestimmten Uni-Abschlusses messen, wenn Studierende über einer gewissen Punktzahl im Eintrittsexamen zugelassen würden.

Daneben gibt es noch einige weitere relevante Methoden wie die «Event Study Difference in Differences»-Analyse oder synthetische Kontrollgruppen. Bei der ersteren Methode wird analysiert, ob sich zwei ähnliche Gruppen nach einer Reform anders entwickeln, wenn eine Gruppe der Reform ausgesetzt ist und die andere nicht. Bei der zweiten wird eine «synthetische» Kontrollgruppe aus einer Kombination anderer Vergleichsgruppen konstruiert. Diese Methode kann man zum Beispiel anwenden, wenn man einen bestimmten Kanton betrachtet und eine Vergleichsgruppe braucht: Die Vergleichsgruppe könnte sich hier aus einer Kombination anderer Kantone zusammensetzen.

«Impact-Award» der Deza

Solche kausalen Evaluationsmethoden haben sich international schnell verbreitet. Viele Verwaltungen haben interne Beratungsstellen geschaffen.² Auch in der Schweiz sind entsprechende Bestrebungen im Gange. So hat die Direktion für Entwicklung und Zusammenarbeit (Deza) einen «Impact-Award» eingeführt, der schweizerische Nichtregierungsorganisationen motiviert, eigene Wirkungsanalysen durchzuführen. Zudem gab sie eine externe Studie in Auftrag, die untersuchte, wie wissenschaftliche Evidenz bei der Deza noch weiter eingebunden und gestärkt werden kann.³

Auch viele Firmen und internationale Organisationen führen Wirkungsanalysen durch, um ihr Vorgehen auf die bestmögliche Evidenz zu basieren. Sie können dabei auch viel von bereits existierenden Studien aus anderen Ländern lernen: Das Forschungsnetzwerk Poverty Action Lab (J-PAL), welches auf die Armutsbekämpfung im weiteren Sinn fokussiert, stellt beispielsweise auf seiner Website Hunderte von Analysen zur Verfügung. Daneben haben sich auch spezialisierte Organisationen entwickelt, die Behörden und Organisationen beim Erarbeiten und Nutzen von Evidenz behilflich sind. Auf internationalem Niveau ist dies etwa Evidence Action und in der Schweiz Policy Analytics.

Alles in allem stellt die neue empirische Revolution im öffentlichen Sektor eine grosse Chance dar. Die neuen Werkzeuge erlauben es den Institutionen, ihre Entscheidungen aufgrund von zunehmend verlässlicheren, evidenzbasierten Grundlagen zu fällen. Es ist zu erwarten, dass sich dieser Trend in den nächsten Jahren noch verstärken wird, wenn sowohl die Verfügbarkeit der Daten wie auch die Expertise in den neuen Methoden weiter zunehmen wird.

- 2 Vgl. Pomeranz und Vila-Belda (2019).
- 3 Roquet et al. (2017).



Dina Pomeranz

Assistenzprofessorin für angewandte Ökonomie am Institut für Volkswirtschaftslehre der Universität Zürich

Literatur

- J-PAL (2015). Why Randomize? Case Study. Cambridge, MA.
- Pomeranz, Dina (2017). Impact Evaluation Methods in Public Economics: A Brief Introduction to Randomized Evaluations and Comparison with Other Methods. In: *Public Finance Review*, 41(1): 10–43.
- Pomeranz, Dina und José Vila-Belda (2019). Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities. *Annual Review of Economics*, 11: 755–781.
- Roquet, Hervé, Bartłomiej Kudrzycki, Adina Rom, Laura Metzger und Isabel Günther (2017). Research Evidence and Impact Evaluations at SDC. ETH Zurich, NADEL – Center for Development and Cooperation.