

Der Verein für Socialpolitik (VfS) – die Vereinigung der deutschsprachigen Ökonomen – hat auf seiner Jahrestagung im September dieses Jahres *Leitlinien zu Standards von Evaluationen wirtschaftspolitischer Maßnahmen* verabschiedet. Diese Leitlinien formulieren einen einheitlichen Bewertungsmaßstab für die Güte von Programmevaluationen, also von empirischen Studien, die die Auswirkungen wirtschaftspolitischer Eingriffe durch den Staat analysieren. Zwei Kernforderungen an staatliche Eingriffe lassen sich nur durch Ex-post-Evaluationen beurteilen: Effektivität (Erreicht die Maßnahme ihr Ziel?) und Effizienz (Wird das Ziel zu möglichst niedrigen Kosten erreicht?). Obwohl kausale Wirkungsanalysen für die Beurteilung staatlicher Eingriffe hinsichtlich dieser zwei Kernanforderungen unumgänglich sind, finden sie insbesondere in der deutschen Innovationspolitik kaum Anwendung. Dass moderne Evaluationsmethoden dennoch auch in diesem Bereich der Wirtschaftspolitik Rückschlüsse darüber erlauben, ob politische Programme für beobachtete Wirkungen tatsächlich ursächlich sind, zeigt die internationale Evidenz.¹ Die Leitlinien des Vereins für Socialpolitik stellen somit einen unerlässlichen Schritt dar, um die Evaluationspraxis der deutschen Politikberatung internationalen Standards anzupassen.

Bereits 2013 bezog der Wissenschaftliche Beirat des Bundesministeriums für Wirtschaft und Energie (BMWi) in einem Gutachten hierzu deutlich Stellung (BMWi 2013): Wirtschaftlichkeitsuntersuchungen sind aufgrund der Bundeshaushaltsordnung nicht nur eine gesetzliche Notwendigkeit (BHO §7 Abs. 2), sondern auch für evidenzbasierte wirtschaftspolitische Entscheidungen unerlässlich. Dennoch seien für nahezu 85% der von Bundesministerien und nachgeordneten Behörden gemeldeten finanzwirksamen Maßnahmen bislang keine Evaluationen durchgeführt worden (vgl. Bundesrechnungshof 2013). Zudem bemängelt der Wissenschaftliche Beirat die Qualität der erstellten Gutachten, da diese »häufig nicht [die] Mindestanforderungen an eine überzeugende Evaluierung wie etwa den Vergleich von geförderten und nicht geförderten Objekten anhand objektiverer Kennzahlen« (BMWi 2013, S. 5) erfüllten. Auf Grundlage solcher Studien lassen sich die ursächlichen Wirkungen der evaluierten Politikmaßnahmen nicht nachweisen. Damit fehlt ihnen auch fast jegliche Aussagekraft für zukünftige wirtschaftspolitische Entscheidungen. Besonderes Augenmerk legt der Beirat auf die unzureichende Qualität von Evaluationsstudien im Bereich der Innovationspolitik. Das Fördervolumen in diesem Politikbereich ist beträchtlich: Allein im Jahr 2013 hat das BMWi über 5 Mrd. Euro zum Zweck der Wirtschaftsförderung bereitgestellt, wobei rund 10% auf das technologieoffene Zentrale Innovationsprogramm Mittelstand (ZIM) entfielen. Gemäß dem Regierungsentwurf für 2014 hat das Bundesministerium für Bildung und Forschung (BMBF) so-

gar einen Etat in Höhe von 5,8 Mrd. Euro für direkte Projektförderungen und Ressortforschung veranschlagt, wovon etwa 57% für Ausgaben im Bereich Forschung und Entwicklung (FuE) eingeplant sind (vgl. BMBF 2014).

Im Folgenden werden zuerst die Grundzüge der neuen Leitlinien des VfS kurz dargestellt und mit der Maryland Scientific Methods Scale (SMS), einem anerkannten internationalen Maßstab, verglichen. Diese Maßstäbe werden dann auf ein praktisches Beispiel der Evaluierung eines deutschen innovationspolitischen Programms angewendet. Abschließend wird anhand von Beispielen aus anderen Politikfeldern in Deutschland sowie aus dem Ausland dargelegt, dass gute Evaluationen durchaus umsetzbar sind und entscheidend zu einer besseren Politik beitragen können.

Leitlinien des VfS: Mindestanforderungen an die Evaluationspraxis

Die Leitlinien des VfS spiegeln den aktuellen wissenschaftlichen Konsens zur Evaluationsmethodik wider und dienen dementsprechend als »Grundlage für die Interpretation, Bewertung und Diskussion der Qualität von Evaluationsstudien« (VfS 2015). Über die Grundvoraussetzungen für Evaluationsstudien hinaus (insbesondere Transparenz, Unabhängigkeit und Unbefangenheit, Nachvollziehbarkeit und Replizierbarkeit, offener Umgang mit Limitationen und Objektivität) werden konkrete Anforderungen an die Wahl der verwendeten Methoden und Daten gestellt.

Die grundlegende Logik von Evaluationen entspricht derjenigen wissenschaftlicher Experimente: Damit die kausale Wirkung einer Maßnahme dargelegt werden kann, müsste eine Untersuchungseinheit (im wirtschaftlichen Kontext z.B. ein Unternehmen, eine Person oder ein Haushalt) jeweils in zwei Zuständen beobachtbar sein: In einem Zustand wirkt die zu evaluierende Maßnahme, im anderen Zustand wurde sie unterlassen (das sogenannte Kontrafaktum). Wäre dies möglich, könnten Unterschiede in den Zielgrößen, anhand derer der Erfolg des Eingriffs beurteilt werden soll, eindeutig auf die Politikmaßnahme zurückgeführt werden. Da eine solche simultane Beobachtung einer Untersuchungseinheit naturgemäß unmöglich ist, erfordern kausale Evaluationsmethoden den Vergleich einer Gruppe von Untersuchungseinheiten, die von der Politikintervention betroffen sind (Interventionsgruppe), mit einer weiteren Gruppe von Untersuchungseinheiten, die nicht betroffen sind (Kontrollgruppe). Die Einheiten in beiden Gruppen sollten bis auf die Inzidenz des Eingriffs in allen weiteren relevanten Eigenschaften ursprünglich – so weit möglich – identisch sein. Nicht gegeben ist dies insbesondere dann, wenn die Aufnahme in die Interventionsgruppe nicht zufällig geschieht. Können beispielsweise nur die forschungstärksten Unternehmen an einer Fördermaßnahme teilnehmen, dann wird

¹ Einen Überblick über evaluierte innovationspolitische Maßnahmen in OECD-Staaten geben die Berichte des von der britischen Regierung eingesetzten »What Works Center for Local Economic Growth« (<http://www.whatworksgrowth.org/policies/innovation/>).

Tab. 1
Evaluationsmethoden sowie ihre Beurteilung auf Basis der Maryland SMS

Stufe	Untersuchungsdesign	Methoden
5	Explizit zufällige Zuteilung von Objekten (z.B. Individuen oder Unternehmen) in Interventions- und Kontrollgruppe (Experimente)	Randomisierte Feldexperimente
4	Quasi-zufällige Zuteilung von Objekten in Interventions- und Kontrollgruppe (Beobachtungsdaten)	Instrumentvariablen, Regression-Discontinuity-Design (RDD)
3	Vergleich der Entwicklung der Zielgrößen in der Interventionsgruppe mit der entsprechenden Entwicklung in einer geeigneten Kontrollgruppe	Differenz-in-Differenzen (DiD), Komplexere Matching-Ansätze
2	Einfacher Vergleich teilnehmender und ausgeschlossener Objekte unter Berücksichtigung beobachtbarer Eigenschaften der Objekte	Querschnittsdaten oder Vorher-Nachher-Vergleich mit Kontrollvariablen, einfache Matching Ansätze
1	Einfacher Vergleich teilnehmender und ausgeschlossener Objekte ohne Berücksichtigung beobachtbarer Eigenschaften der Objekte	Querschnittsdaten oder Vorher-Nachher-Vergleich ohne Kontrollvariablen

Quelle: Zusammenstellung des ifo Instituts.

die Wirkung dieser Politik überschätzt, wenn der durchschnittliche Forschungsoutput dieser Unternehmen mit demjenigen von nicht geförderten (und weniger forschungsstarken) verglichen wird. Bei diesem Vorgehen würde der sogenannte Selektionseffekt vernachlässigt.

Den Empfehlungen des VfS liegt die fundamentale Logik des Kontrollgruppenansatzes zugrunde. Prinzipiell stellt die moderne Ökonometrie viele geeignete Methoden bereit, um einen solchen Kontrollgruppenansatz in der Evaluationspraxis durchzuführen. Allerdings ist die Auswahl der am besten geeigneten Methode entscheidend vom konkreten Untersuchungsgegenstand und der Qualität der zur Verfügung stehenden Daten abhängig. Insbesondere die Schnittstelle von Politik und Wissenschaft erweist sich in Deutschland häufig als problematisch. Aus diesem Grund fügt der VfS seinen generellen Leitlinien für Ex-post-Wirkungsanalysen auch *Empfehlungen an Entscheidungsträger und -trägerinnen in Politik und Verwaltung* an. In diesen stellt der VfS fest, dass für jede zu evaluierende Maßnahme bereits ex ante festgelegt werden muss, anhand welcher (politisch definierten) Zielgrößen und Erfolgskriterien sie ex post zu evaluieren ist. Ebenso sollte die Evaluation auch bereits im Planungs- bzw. Gesetzgebungsverfahren der Maßnahme mit berücksichtigt werden: Nur so kann von Anfang an sichergestellt werden, dass die notwendigen Daten für die Interventions- sowie Kontrollgruppe(n) auch tatsächlich erhoben werden und schlussendlich eine kosteneffiziente Evaluation möglich ist. Mit diesen Empfehlungen verdeutlicht der VfS, dass die Verantwortung für gute Evaluationsstudien, die belastbare Evidenz für die Politikberatung bereitstellen können, gleichermaßen bei Politik und Wissenschaft zu suchen ist.

Ein praktisches Beispiel für die mögliche Umsetzung bietet die Maryland SMS, die die Belastbarkeit infrage kommender Evaluationsmethoden anhand einer Fünfpunkteskala bewertet. Der »Goldstandard« und auf der Skala allein mit fünf Punkten bewertet sind randomisierte Feldexperimente. Vergleichbar zu klinischen Tests werden potenzielle Untersuchungseinheiten hierbei zufällig der Interventions- und der Kontrollgruppe zugewiesen. Tabelle 1 gibt die gängigsten Evaluationsmethoden sowie ihre Beurteilung auf der Maryland SMS wieder.²

Dass diese Skala nicht nur von theoretischem oder akademischem Interesse ist, zeigt ein Blick nach Großbritannien: Im Rahmen des *What Works Network* findet sie dort bereits explizit Anwendung. Das Ziel des Netzwerks ist es, die öffentliche Leistungsbereitstellung effektiver und effizienter

zu gestalten, indem belastbare wissenschaftliche Erkenntnisse besser verbreitet und gezielt für die politische Entscheidungsfindung nutzbar gemacht werden. Als belastbar gilt die Methodik einer Evaluationsstudie erst dann, wenn sie mindestens drei Punkte auf der Skala erreicht. Der Rückgriff auf die Maryland SMS ermöglicht den beteiligten *What Works Centres* hierbei, ihren Bewertungsprozess für die Evaluation verschiedener politischer Maßnahmen transparent zu gestalten und konkrete Erkenntnisse mit praktischer Relevanz für die Wirtschaftspolitik zu gewinnen.

Evaluationsstudien in der deutschen Innovationspolitik: Das Beispiel ZIM

Als nächstes sollen diese Bewertungsmaßstäbe auf ein konkretes Evaluationsbeispiel aus der deutschen Innovationspolitik angewendet werden: das ZIM. Wie oben aufgezeigt, handelt es sich hierbei um ein innovationspolitisches Programm mit erheblicher Budgetrelevanz: Bis heute hat das BMWi im Rahmen des Programms ZIM-Projekte mit einem gesamten Fördervolumen von 4,4 Mrd. Euro bewilligt (vgl. BMWi 2015a).

Als eines der Aushängeschilder der deutschen Innovationspolitik verdient das ZIM besondere Aufmerksamkeit. Das technologie- und branchenoffene Programm wurde im Jahr 2008 vom BMWi mit der Zielstellung ins Leben gerufen, »die Innovationskraft und Wettbewerbsfähigkeit der Unternehmen, einschließlich des Handwerks und der unterneh-

² Eine praktische Anleitung zur Bewertung der Methodik und ihrer Implementierung bietet der Scoring Guide des »What Works Centre for Local Economic Growth« (<http://www.whatworksgrowth.org/public/files/Scoring-Guide.pdf>).

merisch tätigen freien Berufe, nachhaltig [zu] unterstütz[en] und damit einen Beitrag zu deren Wachstum verbunden mit der Schaffung und Sicherung von Arbeitsplätzen [zu leisten]« (BMW 2015b, S. 1). Im April dieses Jahres wurden eine großzügigere Mittelausstattung sowie eine neue Richtlinie verabschiedet, die als wesentliche Neuerungen eine Erweiterung der Zielgruppe des Programms, eine Erhöhung der förderfähigen Kosten und die verstärkte Förderung von Auslandsoperationen beinhalteten. Diese Richtlinie stellt nur die bislang letzte Maßnahme in einer Reihe kontinuierlicher Programmweiterungen dar, die das ZIM seit 2008 erlebt hat. Die Bedeutung des ZIM kann auch anhand der Einschätzung des Bundeswirtschaftsministers erlassen werden, der dem ZIM eine zentrale Rolle bei der Erschließung neuer Trends im Rahmen von »Industrie 4.0« beimisst (BMW 2015c).

Inwiefern das Programm seinen politisch definierten Zielsetzungen gerecht wird (d.h. effektiv ist) und ob es diese auch auf dem bestmöglichen Weg erreicht (d.h. effizient ist), ließe sich mit qualitativ hochwertigen Evaluationsstudien feststellen. Die letzte vom BMW in Auftrag gegebene Evaluation kann keine empirisch belastbaren Aussagen hinsichtlich dieser beiden Kernforderungen liefern. Zwischen 2010 und 2012 abgeschlossene ZIM-Projekte wurden durch das RKW Kompetenzzentrum evaluiert (RKW 2014). Grundlage der Evaluation war eine durch das RKW durchgeführte Befragung aller geförderter Unternehmen, die in den jeweiligen Jahren zumindest ein Projekt beendet hatten. Es fehlt somit der Vergleichsgegenstand, da keine Informationen über Unternehmen vorliegen, die keine Förderung erhielten. Somit lassen sich keine kausalen Effekte anhand objektiver Daten ermitteln; dementsprechend würde diese Studie auf der Maryland SMS in die niedrigste Kategorie (ein Punkt) eingeordnet. Die Methodik lässt Ergebnisse wie das folgende als höchst fraglich erscheinen: »Die Mehrheit der FuE-Projekte wäre ohne ZIM-Förderung nicht durchgeführt worden« (ebd., S. 34). Solche Aussagen befragter Unternehmen sind mit Vorsicht zu behandeln, da beispielsweise strategische Überlegungen von Unternehmen, die an der Fortsetzung des Programms interessiert sind, das Antwortverhalten beeinflussen können.

Im Folgenden wird abschließend gezeigt, dass in anderen Politikfeldern und im Ausland ähnlich komplexe Sachverhalte mit wesentlich aussagekräftigeren Methoden evaluiert worden sind.

Ein Blick auf die Evaluationspraxis in der Innovationspolitik anderer Länder

In Deutschland hat die kausalanalytische Evaluation im vergangenen Jahrzehnt vor allem Einzug in die Arbeitsmarktpolitik gehalten. Insbesondere die institutionalisierte Samm-

lung, Aufbereitung und öffentliche Bereitstellung von Forschungsdaten durch das Institut für Arbeitsmarkt- und Berufsforschung (IAB) hat eine effizienzorientierte Neugestaltung arbeitsmarktpolitischer Instrumente ermöglicht (z.B. der ersatzlose Wegfall von negativ evaluierten Arbeitsbeschaffungsmaßnahmen, vgl. BMW 2013). Der internationale Vergleich zeigt jedoch, dass die Anwendung solcher Untersuchungs- und Bewertungsmethoden nicht auf den Arbeitsmarkt beschränkt bleiben muss.

Eine weltweit bekannte innovationspolitische Maßnahme ist das amerikanische Small Business Innovation Research (SBIR)-Programm. Durch das SBIR-Programm wurden Forschungs- und Entwicklungsaktivitäten kleiner und mittlerer Unternehmen (KMU) mit bislang über 27 Mrd. Dollar gefördert. Wallsten (2000) nutzt einen Instrumentenvariablenansatz, um zu zeigen, dass frühere Evaluationsstudien den positiven Beschäftigungseffekt gegenüber dem wahren (kausalen) Effekt stark überbewerteten. Diese hatten insbesondere Mitnahmeeffekte nicht ausreichend berücksichtigt (d.h. Unternehmen benutzen öffentliche Förderung für Forschung, die sie auch privat finanziert und durchgeführt hätten). Stattdessen stellt die Studie fest, dass jeder öffentlich aufgewendete Dollar private F&E-Ausgaben um denselben Betrag schmälert. Seine Studie wäre auf der Maryland SMS mit vier Punkten zu bewerten und bietet demnach verlässliche, kausale Informationen über die Wirksamkeit dieser politischen Fördermaßnahme unter der Verwendung von objektiven Beobachtungsdaten.

Ebenfalls in den USA angesiedelt sind die Small Business Development Centers (SBDC), die Beratungsdienstleistungen im Bereich der Unternehmensgründung anbieten. Mit Hilfe einer randomisierten Feldevaluierung im Rahmen des »Growing America through Entrepreneurship« (GATE) konnte gezeigt werden, dass die SBDC kurzfristig zwar die Unternehmensgründung erleichtern, die gegründeten Unternehmen aber nicht längerfristig am Markt bestehen (vgl. Fairlie et al. 2015). Die Randomisierung im Untersuchungsdesign (fünf Punkte auf der Maryland SMS) erlaubte hierbei, besonders erfolgreiche Aspekte der Fördermaßnahme zu identifizieren und den politischen Entscheidungsträgern evidenzbasierte Wirtschaftspolitik zu ermöglichen. Zugleich legt dies eindrucksvoll dar, dass auch wirtschaftspolitische Maßnahmen nach höchsten Bewertungsmaßstäben evaluiert werden können.

Ein letztes Beispiel aus den Niederlanden soll noch aufzeigen, dass die Anwendung empirischer Methoden der Kausalanalyse auch systematisch institutionalisiert werden kann. Im Jahr 2004 veranstaltete das niederländische Wirtschaftsministerium (Ministerie van Economische Zaken) erstmalig eine Lotterie über 100 »Innovationsgutscheine«, die zufällig auf die Bewerber aufgeteilt wurden (vgl. Cornet et al. 2006). Das in Den Haag ansässige Centraal Planbu-

reau konnte anschließend in einer Evaluationsstudie zeigen, dass es keine systematischen Unterschiede zwischen den Gewinnern der Innovationsgutscheine und den übrigen Bewerbern gab, wodurch zuverlässig kausale Aussagen hinsichtlich der Wirksamkeit des Programmes gemacht werden konnten.

Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft und Energie (2013), *Evaluierung wirtschaftspolitischer Fördermaßnahmen als Element einer evidenzbasierten Wirtschaftspolitik*, Gutachten des Wissenschaftlichen Beirates beim Bundesministerium für Wirtschaft und Energie, Berlin.

Schlussbemerkungen

Nur durch belastbare Studien, die kausale Effekte nachweisen, lassen sich die Effektivität und Effizienz von Politiken eindeutig belegen. Aufgrund dieser Erkenntnisse kann die Gestaltung und Durchsetzung von Politikmaßnahmen in Zukunft verbessert werden. Diese Beispiele zeigen, dass moderne Methodik in anderen Ländern – anders als in Deutschland – auch in der Innovationspolitik erfolgreiche Anwendung findet. Es ist im allgemeinen Interesse, dass sich diese Entwicklung auch in Deutschland fortsetzt, um besser zu verstehen, welche Maßnahmen tatsächlich wirken. In diesem Sinne sollte die Innovationspolitik in Deutschland ihrem Namen buchstäblich gerecht und zum Vorreiter in der deutschen Wirtschaftspolitik werden, so dass komplexe Problemstellungen mit Hilfe moderner Methoden analysiert und innovative Lösungen gefunden werden können.

Literatur

Bundesministerium für Bildung und Forschung (2014), *Bundesbericht Forschung und Innovation 2014*, Berlin.

Bundesministerium für Wirtschaft und Energie (2015a), *Zentrales Innovationsprogramm Mittelstand*, Statistiken, Stand: 26. Oktober 2015.

Bundesministerium für Wirtschaft und Energie (2015b), Bekanntmachung Neufassung der Richtlinie »Zentrales Innovationsprogramm Mittelstand (ZIM)«, *Bundesanzeiger*, BAnz AT 05.05.2015 B1.

Bundesministerium für Wirtschaft und Energie (2015c), *ZIM-News* Nr. 01-2015.

Cornet, M., B. Vroomen und M. van der Steeg (2006), »Do innovation vouchers help SMEs to cross the bridge towards science?«, CPB Discussion Paper 58, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

Fairlie, R.W., D. Karlan und J. Zinman (2015), »Behind the GATE Experiment: Evidence on Effects of and Rationales for Subsidized Entrepreneurship Training«, *American Economic Journal: Economic Policy* 7(2), 125–161.

RKW Kompetenzzentrum (2014), *Wirksamkeit der geförderten FuE-Projekte des Zentralen Innovationsprogramms Mittelstand (ZIM), Fokus: 2012 abgeschlossene ZIM-SOLO und ZIM-KOOP-Projekte*.

Schriftenreihe des Bundesbeauftragten für Wirtschaftlichkeit in der Verwaltung (2013), *Anforderungen an Wirtschaftlichkeitsuntersuchungen finanzwirksamer Maßnahmen nach § 7 Bundeshaushaltsordnung*, Empfehlungen des Präsidenten des Bundesrechnungshofes als Bundesbeauftragter für Wirtschaftlichkeit in der Verwaltung. Band 18, Berlin

Verein für Socialpolitik, *Leitlinien und Empfehlungen des Vereins für Socialpolitik für Ex post-Wirkungsanalysen*, Fassung: 6. September 2015.

Wallsten, S.J. (2000), »The effects of government-industry R&D programs on private R&D: The case of the Small Business Innovation Research Program«, *RAND Journal of Economics* 31(1), 82–100.