

2019 | Leopoldina-Forum Nr. 1

Evaluierung von Finanzmarktreformen - Lehren aus den Politikfeldern Arbeitsmarkt, Gesundheit und Familie

Herausgegeben von Claudia M. Buch und Regina T. Riphahn



Impressum

Herausgeber der Reihe "Leopoldina-Forum"

Prof. Dr. Dr. h.c.mult. Jörg Hacker

Präsident der Deutschen Akademie der Naturforscher Leopoldina e. V.

– Nationale Akademie der Wissenschaften –

Jägerberg 1

06108 Halle (Saale)

Redaktion

Dr. Katja Patzwaldt, Esther Meyer

Nationale Akademie der Wissenschaften Leopoldina

Kontakt

katja.patzwaldt@leopoldina.org

Redaktionsschluss: November 2018

Gestaltung und Satz

unicom Werbeagentur GmbH, Berlin

Druck

Druckhaus Köthen GmbH & Co. KG, Köthen (Anhalt)

ISBN 978-3-8047-3939-0

Bibliografische Information der deutschen Nationalbibliothek

Die deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie, detaillierte bibliografische Daten sind im Internet unter <http://dnb.d-nb.de> abrufbar.

Zitiervorschlag

Buch CM, Riphahn RT (Hrsg.) (2019): Evaluierung von Finanzmarktreformen - Lehren aus den Politikfeldern Arbeitsmarkt, Gesundheit und Familie. Leopoldina-Forum Nr. 1. Nationale Akademie der Wissenschaften Leopoldina, Halle (Saale).

Evaluierung von Finanzmarktreformen - Lehren aus den Politikfeldern Arbeitsmarkt, Gesundheit und Familie

Herausgegeben von Prof. Dr. Claudia M. Buch
(Deutsche Bundesbank) und Prof. Regina T. Riphahn, PhD
(Nationale Akademie der Wissenschaften Leopoldina)

Vorwort

Evidenzbasierte Politikberatung – Wissenschaft in der politischen Gestaltung stärken

Von JÖRG HACKER

*Präsident, Deutsche Akademie der Naturforscher Leopoldina –
Nationale Akademie der Wissenschaften*

Die Finanzkrise wurde als Immobilien-, Schulden- und Eurokrise in vielen Facetten diskutiert. Sie hat zu hohen Kosten für viele Gesellschaften geführt. Allein in Deutschland betrugen die krisenbedingten Zahlungen zur Stützung der Banken 70 bis 80 Milliarden Euro.¹

Vor diesem Hintergrund ist es wichtig, dass die Wissenschaft verstärkt vertrauenswürdige Informationen und transparente Bewertungen über die Wirkungen der seit der Krise beschlossenen Regulierungen der Finanzmärkte in die öffentliche Meinungsbildung und demokratische Entscheidungsfindung einbringt. In Bezug auf die Wirkung von Regulierungen der Finanzmärkte besteht ein besonderes Transparenz- und Informationsdefizit, wie u.a. der Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung feststellte.²

Unter der deutschen Präsidentschaft der G20 im vergangenen Jahr wurde daher ein Rahmenwerk beschlossen, mit dem die Reformen der Finanzmärkte evaluiert werden können.

Wenn wir davon ausgehen, dass Forschung zu Ursachen und Folgen der Finanzkrise hilfreiches Wissen für die Gesellschaft bereitstellt, müssen wir gerade mit Blick auf die Regulierung der Finanzmärkte fragen, warum wissenschaftliche Evaluierung dort bislang wenig stattfindet und kaum politische Wirkungen erzielt.

Beispiele für wissenschaftliche Evaluierung von politischen Inter-

1 HELLOWIG 2017.

2 SVR 2008; SVR 2007.

ventionen finden sich in der Arbeitsmarkt-, Familien- oder Gesundheitspolitik. Expertinnen und Experten aus diesen Feldern diskutieren in diesem Band, was wir aus diesen Erfahrungen lernen können. Dabei geht es weniger um wissenschaftliche Methoden als um die Frage, wie gute Evaluierungen gelingen können, welche Voraussetzungen dafür im politischen Prozess und der Wissenschaft nötig sind – und wie ein besserer Dialog zwischen beiden Seiten gestaltet werden kann.

Evidenzbasierte Politikberatung generell und gerade auch im Bereich der Finanzmärkte und der wirtschaftspolitischen Beratung insgesamt zu stärken, ist ein Anliegen, das die Leopoldina als Nationale Akademie der Wissenschaften unterstützt. Wissenschaftsgeleitete Evaluierungen sind als Teilbereich von Politikberatung und auch von Wissenschaftskommunikation mit am stärksten auf einen konkreten Handlungskontext ausgerichtet, nämlich die Bewertung von politischen Programmen und Maßnahmen durch Entscheidungsträgerinnen und -träger.

Wir tragen Verantwortung für die Validität und Korrektheit der wissenschaftlichen Stellungnahmen, die wir für die Politikberatung erarbeiten. Unabhängige Forschung ist aber nicht immer so nah an der politischen Gestaltung oder vermag Komplexität und Dynamik gesellschaftlicher Entwicklungen so zu bearbeiten, dass sie unverzüglich mit Analysen und Lösungsvorschlägen aufwarten kann. Daher erscheint es umso wichtiger, wissenschaftliche Evaluierung von politischen Maßnahmen vom Moment ihrer Konzipierung an vorzusehen. Eine gut ausgebaute Infrastruktur für die Erhebung von Daten, die Verpflichtung der Akteure auf wissenschaftliche Auswertung von politischen Interventionen und die klare, offene Kommunikation der Ergebnisse, Grenzen und Möglichkeiten der Evaluierungen können zu guter demokratischer Gestaltung und politischen Debatten beitragen.

Literatur

Hellwig, M. (2017). „Deutschland und die Finanzkrise(n)“. Wirtschaftsdienst 9, 606-607.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (SVR) (2008). „Jahresgutachten 2008/09“. Wiesbaden.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (SVR) (2007). „Jahresgutachten 2007/08“. Wiesbaden.

Inhalt

1. Einführung 7

BUCH, CLAUDIA M. und RIPHAHN, REGINA T.:

Evidenzbasierte Politik: Infrastrukturen für bessere

Informationen 7

LUDEWIG, JOHANNES:

Regulierung und Evaluierung –

Aufgaben des Normenkontrollrats..... 14

2. Politikfeld Arbeitsmarkt 17

SCHMIDT, CHRISTOPH M.:

Geht doch: Zur Evaluation großer Reformpakete am

Beispiel der Arbeitsmarktpolitik 17

WALWEI, ULRICH:

Evidenzbasierte Beratung im Politikfeld Arbeitsmarkt 34

BÖRSCH-SUPAN, AXEL:

Evidenzbasierte Beratung in der Rentenpolitik..... 47

FITZENBERGER, BERND:

Evidenzbasierte Wirtschaftspolitik:

Lehren aus dem Politikfeld Arbeitsmarkt..... 58

3. Politikfeld Finanzmärkte und Fiskalpolitik..... 67

FREIHUBE, THORSTEN und HUFELD, FELIX:

Evidenzbasierte Finanzmarktregulierung –

Implikationen aus Sicht der Finanzaufsicht..... 67

ROCHOLL, JÖRG:

Evidenzbasierte Evaluierung als Voraussetzung

für Reformen der Finanzmärkte – Ursachen und

Schlussfolgerungen aus der Finanzkrise 2008..... 75

EBERT, WERNER; DÖNNEBRINK, ELMAR und

SCHUKNECHT, LUDGER:

Gute Evidenz für wirksame Fiskalpolitik 83

4. Politikfeld Gesundheit..... 95

WINTER, JOACHIM:

Evidenzbasierung in der Medizin:

Status quo und neue Herausforderungen..... 95

LEIDL, REINER:

Evidenzbasierung in Medizin, Gesundheitspolitik und

-versorgung: Drei Thesen..... 100

SCHULENBURG, J.-MATTHIAS GRAF VON DER:

Evaluierung von Finanzmarktreformen:

Lehren aus anderen Politikbereichen - Beispiel Gesundheit... 108

5. Politikfeld Familie 121

RISTAU, MALTE:

Evidenzbasierung in der Familienpolitik: Wie sie möglich
wurde, was sie ausmacht und warum Promotoren
notwendig sind 121

MACKROTH, PETRA:

Evidenzbasierung in der Familienpolitik des Bundes 133

SPIß, C. KATHARINA:

Zehn Thesen für eine erfolgreiche evidenzbasierte
Politikberatung: Das Beispiel der Familienpolitik..... 138

BONIN, HOLGER:

Die Gesamtevaluation ehe- und familienbezogener
Leistungen: Meilenstein auf dem Weg zu einer
evidenzbasierten Familienpolitik in Deutschland 149

BERTRAM, HANS:

Zeit, Geld und Infrastruktur: Die neue Familienpolitik -
Begründung und Evaluation..... 156

6. Anhang 171

Programm des Workshops vom 29. Mai 2018 171

1. Einführung

Evidenzbasierte Politik: Infrastrukturen für bessere Informationen¹

Von CLAUDIA M. BUCH und REGINA T. RIPHAHN

Korrespondenzadressen:

Claudia M. Buch
Deutsche Bundesbank

Wilhelm-Epstein-Straße 14
60431 Frankfurt a.M.
office-buch@bundesbank.de

Regina T. Riphahn
Friedrich-Alexander-Universität
Erlangen-Nürnberg
Fakultät für Wirtschaftswissenschaften
Lange Gasse 20, 90403 Nürnberg
regina.riphahn@fau.de

In der Bildungspolitik, in Bezug auf Arbeits- oder Finanzmärkte oder auch bei wohnungspolitischen Maßnahmen – in vielen Politikbereichen müssen Entscheidungen getroffen werden, die auf Evidenz über ihre Wirkung aufbauen sollten. Genauso benötigt die Öffentlichkeit in einem demokratischen Gemeinwesen Informationen darüber, ob gewählte Repräsentantinnen und Repräsentanten ihre politischen Entscheidungen nachvollziehbar in Hinblick auf konkrete Ziele treffen. Die Politik ist, ebenso wie Bürgerinnen und Bürger, an der Frage interessiert, wie öffentliche Mittel und Regulierungen effektiv eingesetzt werden können, um gegebene Ziele zu erreichen. Welche Maßnahmen verbessern die Funktionsweise der Arbeitsmärkte? Warum schneiden manche Bundesländer beim Vergleich von Schulleistungen besser ab als andere? Wie wirken Maßnahmen, mit denen bezahlbarer Wohnraum geschaffen werden soll? Wie wirkt die Regulierung der Finanzmärkte?

1 Wir danken Katja Patzwaldt und Edgar Vogel für sehr hilfreiche Anregungen und Kommentare zu einer früheren Version. Alle Fehler und Ungenauigkeiten liegen allein in unserer Verantwortung.

Die systematische Evaluierung von Politikmaßnahmen kann helfen, solche Fragen zu beantworten. Allerdings ist eine solche evidenzbasierte Politik in Deutschland nur schwach verankert; Evaluierungen sind nicht durchgängig in das Handeln von Verwaltungen und Politik integriert. Das erschwert es, transparente und objektive Diskussionen zu führen.

Dabei hat die Wissenschaft erhebliche Fortschritte bei der Entwicklung von Methoden zur Evaluierung gemacht. Daher kommt es jetzt darauf an, einen konstruktiven Dialog zwischen Wissenschaft und Politik zu stärken und die für Evaluierungen notwendigen Infrastrukturen zu verbessern. Das heißt, Strukturen für Evaluierungen zu schaffen, Informationen über Evaluierungen und deren Ergebnisse leichter öffentlich zugänglich zu machen, Datenzugang zu verbessern und den Transfer wissenschaftlicher Erkenntnisse in die Praxis der Handelnden, aber auch der Evaluierenden zu erleichtern.

Informationsbasis für politische Entscheidungen verbessern

Evidenzbasierte Politik folgt einem kontinuierlichen, faktenbasierten Lernprozess: Sie definiert Ziele, beschließt Maßnahmen und setzt sie um, prüft deren Wirksamkeit, passt Beschlüsse an und lernt für zukünftiges Handeln. So kann Politik effektiver werden: Maßnahmen, die nicht wirken und unerwünschte Nebenwirkungen haben, werden zurückgenommen. Erfolgreiche Konzepte werden gestärkt. Bei der Evaluierung von Maßnahmen sollte die Aussagekraft unterschiedlicher Quellen und Analysemethoden berücksichtigt werden. Eine Vielfalt an Ansätzen – neben systematischen Literaturüberblicken auch empirische Metaanalysen, Längsschnitterhebungen oder Fallstudien – ist wünschenswert. Die zugrundeliegenden Daten sollten für die Forschung zugänglich sein, Ergebnisse veröffentlicht werden und so Qualitätssicherung sowie öffentliche Debatte ermöglichen.

Einem solchen idealtypischen Verlauf der Politikgestaltung und -evaluierung stehen in der Praxis Hindernisse entgegen: Politisches Handeln unterliegt dem Druck, schnelle Ergebnisse zu liefern, so dass für eine ausgewogene Bewertung von Maßnahmen oft keine Zeit ist. Allgemein formulierte Ziele in konkrete Indikatoren zu übersetzen kann schwierig sein. Verwaltungsprozesse, Mitbestimmungswege und andere prozedurale Anforderungen sind einzuhalten und geeignete Daten sind nicht immer vorhanden. Politik ist daher oft eher die „Kunst des Möglichen“ als

das Ergebnis eines auf Evidenz basierenden Prozesses. Nicht zuletzt sind auch auf Seiten der Wissenschaft die Anreize nicht immer so gesetzt, dass wissenschaftliche Transferprojekte und Politikevaluierungen gefördert werden. Dabei gäbe es viel zu gewinnen, wenn politische Entscheidungen verstärkt auf Basis besserer Informationen getroffen würden: Die Maßnahmen wären effektiver und kostengünstiger.

In den vergangenen Jahren sind erhebliche Fortschritte gemacht worden: Nicht nur die Wissenschaft stellt bessere Methoden bereit,² mit denen die Wirksamkeit politischen Handelns untersucht werden kann und vermittelt sie in Studiengängen. Auch in Politik und Verwaltung gibt es vielversprechende Initiativen, diese Erkenntnisse zu nutzen. Nicht zuletzt sind mehr und bessere Daten verfügbar.³

Evidenzbasierte Politik – aktueller Handlungsbedarf

Dem Ziel, politische Maßnahmen besser auf ihre Wirksamkeit hin zu überprüfen, kommt gerade angesichts der aktuellen Debatte über globale Herausforderungen und der Rolle von Expertinnen und Experten in der öffentlichen Diskussion große Bedeutung zu.

Die Reformen der internationalen Finanzmärkte sind hierfür ein augenfälliges Beispiel. Das Ziel der im Nachgang der Finanzkrise beschlossenen Reformen war es, die Widerstandsfähigkeit des Finanzsystems zu stärken und Kosten von möglichen Finanzkrisen den Verursachern aufzuerlegen. In den letzten zehn Jahren wurden viele dieser Reformen umgesetzt. Nun muss evaluiert werden, ob und wie diese Reformen ihre Wirkung entfalten, welche Nebenwirkungen sie haben, welcher

2 Gängige Lehrbücher für Evaluierungsmethoden sind beispielsweise ANGRIST, PISCHKE 2009; ANGRIST, PISCHKE 2015 oder BAUER, FERTIG, SCHMIDT 2009.

3 So setzt sich der Rat für Sozial- und Wirtschaftsdaten für Datenzugang ein und hat über 30 Forschungsdatenzentren akkreditiert (www.ratswd.de). Im Januar 2013 hat der Staatssekretärsausschuss zum Bürokratieabbau die Einführung eines systematischen Evaluierungsverfahrens für alle Gesetze und Verordnungen ab einem Schwellenwert von 1 Mio. Euro jährlichem Erfüllungsaufwand beschlossen. Die Empfehlung des Ausschusses für Finanzstabilität zur Schaffung makroprudenzieller Instrumente im Bereich des Wohnimmobilienmarktes enthält eine Klausel zur Evaluierung der Maßnahmen (AFS 2015). Der Wissenschaftliche Beirat beim Bundeswirtschaftsministerium hat sich in einem Gutachten für die Evidenzbasierung von Politikmaßnahmen ausgesprochen (BEIRAT 2013). Ein Beispiel für systematische Evidenzbasierung findet sich für die Familienpolitik bei BÖHMER ET AL. 2014 (vgl. Beitrag von Spieß in diesem Band) und für die Arbeitsmarktpolitik bei FERTIG ET AL. 2004) (vgl. Beiträge von Fitzenberger und Schmidt in diesem Band). Für einen Überblick zu Abschätzung von Kostenfolgen und Nutzen von Regulierung vgl. NATIONALER NORMENKONTROLLRAT 2018.

Nutzen und welche Kosten für die Gesellschaft insgesamt entstehen. Unter deutscher G20-Präsidentschaft wurde im Jahr 2017 vom Financial Stability Board (2017) ein Rahmenwerk entwickelt, das eine strukturierte Evaluierung von Reformen des Finanzsektors ermöglicht. Dieses Rahmenwerk steht in der Tradition anderer Politik- und Wissenschaftsbereiche, in denen Evaluierungen gängige Praxis sind. Mit der Evaluierung von Finanzmarktreformen gibt es noch wenig Erfahrung. Daher ist zu klären, was sie von anderen Politikbereichen lernen kann.

Die Deutsche Bundesbank und die Leopoldina wollen einen Erfahrungsaustausch sowohl über wissenschaftliche Disziplinen hinweg als auch zwischen Politik und Wissenschaft ermöglichen. Wissenschaftlerinnen und Wissenschaftler sowie Vertreterinnen und Vertreter aus Politik bzw. Verwaltung wurden daher im Mai 2018 zu einem Workshop eingeladen. Dort wurde diskutiert, welche Erfahrungen in den Politikfeldern Arbeitsmarkt, Familie, Gesundheitssystem und Finanzmärkte mit der Evaluierung von Maßnahmen gemacht wurden und welche Faktoren einen guten Dialog zwischen Politik und Wissenschaft bei der Evaluation von Regulierungsvorhaben befördern können.

Dialog zwischen Wissenschaft und Politik intensivieren

Die Diskussion hat gezeigt: Wissenschaftlich fundierte Analysen können die Basis für besser informierte politische Entscheidungen liefern, politische Entscheidungsprozesse bereichern und damit letztlich deren Legitimität erhöhen. Für eine stärker evidenzbasierte Politik sind folgende Aspekte entscheidend:

Erstens müssen Politik, Verwaltung und die Öffentlichkeit überzeugt sein, dass evidenzbasierte Entscheidungen einen Mehrwert bringen. Dies umfasst Akteure im jeweiligen Politikfeld, von bestimmten Maßnahmen Betroffene und auch die Medien. Dabei können die Akteure ganz unterschiedliche Funktionen haben: Ärztinnen und Ärzte müssen beispielsweise gesundheitspolitische Leitlinien umsetzen, Lehrerinnen und Lehrer sollten auf Basis von Vergleichen der Leistungen von Schulen und Klassen handeln, auf dem Arbeitsmarkt sind Bildungsträger oder Personalverantwortliche gefragt. Bei positiver Evaluierung können Maßnahmen fortgesetzt werden. Bei negativer Evaluierung ergeben sich Ansatzpunkte für Verbesserungen. Damit der Weg für methodisch angemessene Evaluationsstudien geöffnet ist, müssen zudem Rechts-

grundlagen geschaffen werden, Ziele definiert und frühzeitig Indikatoren der Zielerreichung festgelegt werden. Eine gesetzlich verankerte Pflicht zur Evaluierung beugt Interessenskonflikten vor und überwindet oben genannte Hindernisse in der Praxis.

Zweitens sollte der Zugang zu relevanten Informationen erleichtert werden, beispielsweise über digitale Plattformen. Auf solchen Plattformen können Informationen über unterschiedliche Studien zur Verfügung gestellt werden, aber auch Informationen über relevante Regulierungen und Institutionen. So kann der Dialog zwischen Wissenschaft, Politik und Öffentlichkeit verbessert, eine verständliche Kommunikation von wissenschaftlichen Ergebnissen, Methoden, ihren Möglichkeiten und Einschränkungen ermöglicht und Transparenz über Rahmenbedingungen in der Verwaltung verbessert werden. Zugleich ist es erforderlich, die Belastbarkeit von empirischer Evidenz, die mitunter nur von Expertinnen und Experten einzuschätzen ist, in verständlicher Weise darzustellen.

Drittens sollten notwendige Daten verfügbar sein – zur Nutzung innerhalb der Verwaltungen, aber auch für externe Forscherinnen und Forscher. Daten im Nachhinein zu erheben kann sehr teuer oder sogar unmöglich sein. Daher sollten die für die Analyse erforderlichen Daten bereits vor Umsetzung einer Maßnahme bestimmt werden. Pilotstudien können dabei helfen, den Datenbedarf zu ermitteln. Die Einhaltung des Datenschutzes ist dabei zentral. Je mehr Personen mit den Daten arbeiten und aus den Erfahrungen anderer lernen, desto höher ist langfristig die Qualität von Evaluationsstudien.

Viertens sollten im wissenschaftlichen Bereich die Anreize verbessert werden, politische Entscheidungsprozesse durch gute Evidenz zu unterstützen. Transferorientierte wissenschaftliche Zeitschriften, wie im Gesundheits- und Bildungswesen oder Publikationen von Replikationsstudien,⁴ akademische Preise und Forschungsförderung in diesem Bereich, wie etwa das Forschungsnetzwerk Alterssicherung (FNA) der Deutschen Rentenversicherung, sowie das Engagement von wissenschaftlichen Fachgesellschaften in der Bewertung von Evidenz können solche Anrei-

4 Diese existieren etwa im Gesundheits- und Bildungswesen mit den Zeitschriften *Evaluation & The Health Professions* oder *Studies in Educational Evaluation*. Das *International Journal for Reviews in Empirical Economics* veröffentlicht explizit Replikationsstudien.

ze vermitteln. Gerade Deutschland hat mit zahlreichen unabhängigen Forschungsinstituten und Forschungsdatenzentren zudem eine hervorragende Infrastruktur, die noch besser als bisher für einen Transfer wissenschaftlicher Erkenntnisse in die Praxis genutzt werden kann. Wissenschaftliche Forschung ist insbesondere erforderlich, um Kausalzusammenhänge zu untersuchen und Erklärungen, die in konkreten Evaluierungen nicht notwendigerweise eine Rolle spielen, zu identifizieren. Das ist beispielsweise erforderlich, wenn es um spezifische Kontextfaktoren oder die Abwägung von Kosten und Nutzen eines bestimmten Programms geht.

„Mehr Evidenz“ in politische Entscheidungsprozesse zu bringen, ist ein Ziel im Interesse der Gesellschaft. Um auf diesem Weg weiter zu kommen, sind Offenheit und Dialogbereitschaft von Wissenschaft und Politik erforderlich. Je besser die vorhandenen Strukturen genutzt werden, je besser Evaluierungen in laufende Prozesse eingebettet werden, umso geringer werden die Kosten von Evaluierungen – und umso mehr kann die Gesellschaft aus Erfahrungen lernen.

Literatur

Angrist, J.D., Pischke, J.-S. (2009). „Mostly Harmless Econometrics: An Empiricist’s Companion“. Princeton: Princeton University Press.

Angrist, J.D., Pischke, J.-S. (2015). „Mastering ‘Metrics: The Path from Cause to Effect“. Princeton: Princeton University Press.

AFS (Ausschuss für Finanzstabilität) (2015). „Empfehlung vom 30. Juni 2015 zu neuen Instrumenten für die Regulierung der Darlehensvergabe zum Bau oder Erwerb von Wohnimmobilien“. Berlin.

Bauer, T.K., Fertig, M., Schmidt, C.M. (2009). „Empirische Wirtschaftsforschung: Eine Einführung“. Springer.

Beirat (Wissenschaftlicher Beirat beim Bundesministerium für Wirtschaft und Energie) (2013). „Evaluierung wirtschaftspolitischer Fördermaßnahmen als Element einer evidenzbasierten Wirtschaftspolitik“. Bundesministerium für Wirtschaft und Energie. Berlin.

Böhmer, M., Ehrentraut, O., Heimer, A., Henkel, M., Ohlmeier, N., Poschmann, K., Schmutz, S., Weisse, J. (2014). „Gesamtevaluation der ehe- und familienbezogenen Maßnahmen und Leistungen in Deutschland. Endbericht“. Basel. <https://www.bmfsfj.de/blob/93954/25490622c47497e47acbcfa797748cfb/gesamtevaluation-der-ehe-und-familienbezoge->

nen-massnahmen-und-leistungen-data.pdf

[Zugriff 7. Juli 2018]

Bundesministerium der Finanzen (2017). „Forschungsvorhaben aus: fe 5/17: Evaluierung gesamt- und finanzwirtschaftlicher Effekte der Reformen europäischer Finanzmarktregulierung im deutschen Finanzsektor seit der Finanzkrise“. Berlin.

Fertig, M., Kluve, J., Schmidt, C.M., Apel, H., Friedrich, W., Hägele, H. (2004). „Die Hartz-Gesetze zur Arbeitsmarktpolitik: Ein umfassendes Evaluationskonzept“. Berlin: Duncker & Humblot.

Financial Stability Board (2017). „Framework for Post-Implementation Evaluation of the Effects of the G20 Financial Regulatory Reforms“. Basel.

Nationaler Normenkontrollrat (2018). „Gesamtkonzept Bürokratieabbau und bessere Rechtsetzung“. https://www.normenkontrollrat.bund.de/Webs/NKR/DE/UeberUns/Gesamtkonzept/_node.html

[Zugriff 11.09.2018]

Regulierung und Evaluierung – Aufgaben des Normenkontrollrats

Von JOHANNES LUDEWIG

Korrespondenzadresse:
Nationaler Normenkontrollrat
Bundeskanzleramt
11012 Berlin
nkr@bk.bund.de

Mit seiner Einsetzung im Jahr 2006 erhielt der Nationale Normenkontrollrat den Auftrag, die Bundesregierung auf dem Gebiet der besseren Rechtssetzung als unabhängiger Berater zu unterstützen. Gesetze sollen die beabsichtigten Wirkungen haben, können aber auch unbeabsichtigte Nebenfolgen auslösen. Eine gute Evaluierung nach angemessener Zeit macht sichtbar, ob die Ziele erreicht wurden und ob Nebenfolgen eingetreten sind. Dadurch ermöglicht sie dem Gesetzgeber eine Nachsteuerung und kann zugleich die Akzeptanz der Regelung bei Bürgerinnen und Bürgern, Wirtschaft und Verwaltung erhöhen.

Im März 2013 hat die Bundesregierung deshalb - ausgehend von einem Vorschlag des Normenkontrollrats - die „Konzeption zur Evaluierung neuer Regelungsvorhaben“ beschlossen. Diese Konzeption bildet seitdem die Grundlage für Evaluierungen der Bundesministerien. Danach ist ein Regelungsvorhaben nach drei bis fünf Jahren zu evaluieren, wenn der jährliche Erfüllungsaufwand mehr als eine Million Euro beträgt. Anlass für eine Evaluierung kann auch der hohe finanzielle Gesamtaufwand, die besondere politische Bedeutung oder große Unsicherheit über die Wirkungen eines Regelungsvorhabens sein. In der Konzeption fehlen jedoch einheitliche Vorgaben zur Methodik der Evaluierung selbst, so dass jedes Ressort in eigener Zuständigkeit entscheidet, wie es hier vorgeht – ein Umstand, der Qualität und Vergleichbarkeit nicht gerade fördert.

Der Nationale Normenkontrollrat prüft die Erwägungen zur späteren Evaluierung in Gesetzes- und Verordnungsentwürfen und überwacht die Einhaltung der Evaluierungspflicht. Damit ist er *ex ante* sowie *ex post* in die Evaluierungsverfahren der Bundesregierung eingebunden.

Schon im Gesetzgebungsverfahren achtet der Normenkontrollrat darauf, dass die Evaluierung im sogenannten Dreiklang angelegt wird: Dafür sind erstens die zu evaluierenden Ziele, zweitens die Kriterien/Indikatoren für die Messung der Zielerreichung und drittens die dafür notwendigen Datenquellen zu definieren. Wichtigstes Evaluierungskriterium ist die Zielerreichung, da Regelungen, die das im Regelungsentwurf formulierte Ziel verfehlen, meist unnötigen Erfüllungsaufwand verursachen. Weitere Indikatoren oder Messgrößen müssen hinzukommen, damit Fragen formuliert werden, deren Beantwortung zu einer Bewertung führen kann. Die Vorbereitung der Evaluierung muss auch klären, wie die für eine gute Evaluierung notwendigen Daten bereits während der Anwendung des Gesetzes beschafft bzw. erhoben werden können.

Eine anschließende Qualitätskontrolle rundet das Verfahren ab. Gegenwärtig gibt es allerdings noch keine regierungsinterne oder externe Stelle mit einem Mandat, die Qualität der Evaluierungen zu überwachen. Es fehlt zudem ein Verfahren, um betroffene Verbände oder andere Behörden zu beteiligen, so dass wertvolles Praktikerwissen unberücksichtigt bleibt.

Der Staatssekretärsbeschluss zur Evaluierung war zweifellos ein wichtiger Schritt zur Stärkung der Evidenzbasierung in der Bundesregierung. Evaluierungen werden dadurch zunehmend Teil der alltäglichen Praxis der Bundesverwaltung. Um die Ergebnisse zu verbessern und den Nutzen zu steigern, muss das aktuelle Verfahren weiterentwickelt werden. Mit verbindlichen Mindeststandards, klarer Zuständigkeit für Qualitätskontrolle sowie regelmäßiger Beteiligung von Dritten können Evaluierungsergebnisse und ihre Akzeptanz verbessert werden. Der Normenkontrollrat hat im Jahr 2017 ein Phasenmodell für eine gute Evaluierung vorgelegt und den Ressorts zur Verfügung gestellt. Damit gibt es bereits einen konkreten Vorschlag, wie einheitliche Mindeststandards bei Evaluierungen umgesetzt werden können.

Eine aus methodischer Sicht vorbildliche Evaluierung kann ihren Beitrag zur besseren Rechtssetzung letztlich nur dann leisten, wenn die Ergebnisse bei zukünftigen Entscheidungen berücksichtigt werden. Klare Handlungsempfehlungen am Ende jedes Evaluationsberichts sind dazu unbedingt notwendig. Geschlossen wird der Rechtssetzungskreislauf durch eine systematische Verknüpfung zwischen den Ergebnissen einer

Evaluierung und einem neuen Vorhaben. Gesetze oder Verordnungen sollten deshalb nur unter Berücksichtigung einer vorangegangenen Evaluierung geändert werden – nach dem Motto „keine Novellierung ohne Evaluierung“ - ein Grundsatz, den die EU-Kommission bereits eingeführt hat. Eine sachgerechte Evaluierung ist und bleibt damit zentrales Element evidenzbasierter politischer Vorhabenplanung.

2. Politikfeld Arbeitsmarkt

Geht doch: Zur Evaluation großer Reformpakete am Beispiel der Arbeitsmarktpolitik¹

Von CHRISTOPH M. SCHMIDT

Korrespondenzadresse:

RWI – Leibniz-Institut für Wirtschaftsforschung

Postfach 10 30 54

45030 Essen

praesident@rwi-essen.de

1. Evaluation umfassender Reformpakete: Der Präzedenzfall

Akute Krisen und verfestigte Problemlagen machen umfassende Reformen notwendig. Das herausragende Beispiel der jüngeren Vergangenheit ist die Finanzkrise der Jahre 2008/09: Sie hat aufgedeckt, dass es einer umfassenden Reform des Regelwerks, der Institutionen und Prozesse bedarf, um die Stabilität der Finanzmärkte zu gewährleisten. Doch hier gilt wie in allen anderen Bereichen wirtschaftspolitischen Handelns, dass ein gutes Vorhaben schon allein angesichts der zu überwindenden Widrigkeiten nicht die erfolgreiche Tat ersetzt. Ein derart umfassendes Reformpaket sollte vielmehr im Hinblick auf die Wirksamkeit einzelner Maßnahmen und des Gesamtpakets sorgfältig evaluiert werden.²

Nun ist die Wirkung eines Reformpakets, das alle Elemente des Systems berührt und alle Akteure betrifft und dessen vielfältige Elemente einander überlagern, zweifelsfrei alles andere als leicht zu erfassen. Der vorliegende Beitrag zeigt jedoch, dass es gute Gründe für Optimismus

1 Ich bedanke mich für konstruktive Anregungen bei Thomas Bauer, Andrea Cassel, Jochen Kluve, Wim Kösters, Wolf Reuter und Astrid Schürmann. In besonderem Maße habe ich bei diesem Beitrag auf frühere gemeinsame Arbeiten mit Michael Fertig zurückgegriffen, dem ich in Dankbarkeit gedenke.

2 DEUTSCHE BUNDESBANK 2017; FSB 2017.

gibt. Dabei dient die Evaluation der Anfang des Jahrhunderts durchgeführten umfassenden Hartz-Reformen des Arbeitsmarkts als erfolgreicher Präzedenzfall. Die mit diesem Reformpaket verbundenen Änderungen zielten darauf ab, die Vermittlung in Arbeit rascher und nachhaltiger zu gestalten, neue Beschäftigungsfelder zu schaffen und mehr Brücken in die Beschäftigung zu bauen. Ihr übergreifendes Ziel war letztlich die Verringerung der Arbeitslosigkeit.

Seit dem Jahr 2005 hat der deutsche Arbeitsmarkt eine derart starke Entwicklung hin zum Besseren erlebt, dass nicht wenige Beobachter von einem „Arbeitsmarktwunder“ sprechen: Trotz des starken wirtschaftlichen Einbruchs des Jahres 2009 wurde die Anzahl der Arbeitslosen ungefähr halbiert, die Beschäftigung um mehr als vier Millionen Personen ausgeweitet.³ Der exakte Beitrag, den die Hartz-Reformen dazu geleistet haben, wird sich zwar nie mit Sicherheit bemessen lassen. Doch wir wissen heute weit besser als zuvor, welche Weichenstellungen und Eingriffe dazu beitragen können, den Arbeitsmarkt in ein besseres Gleichgewicht mit hoher Beschäftigung und niedriger Arbeitslosigkeit zu führen.

Diesen Erkenntnisfortschritt über die tatsächlichen kausalen Wirkungen arbeitsmarktpolitischer Maßnahmen und die damit verbundenen Chancen für eine bessere Arbeitsmarktpolitik verdanken wir vor allem den umfassenden Anstrengungen zur Evaluation des Reformpakets. Die Evaluation von Maßnahmen der Arbeitsmarktpolitik erlebte in den Jahren nach den Hartz-Reformen eine regelrechte Hochphase. Ihr Auslöser war allerdings vor allem das Resultat öffentlichen Rechtfertigungsdrucks, nicht etwa ein sprunghafter Anstieg der Lust auf Erkenntnisgewinn: Das einschneidende Ereignis war der Skandal um die teilweise geschönten Vermittlungsstatistiken der damaligen Bundesanstalt für Arbeit Anfang des Jahrhunderts.

Dieser Skandal entstand vor dem Hintergrund eines lang anhaltenden Siechtums der deutschen Volkswirtschaft, das sich insbesondere am Arbeitsmarkt in einer hohen Anzahl (Langzeit-)Arbeitsloser widerspiegelte. Aus dem Skandal erwuchs in der Öffentlichkeit ein großes Misstrauen gegenüber den damit nicht direkt zusammenhängenden Erfolgsstatistiken der aktiven Arbeitsmarktpolitik. Nichtsdestoweniger waren diese Erfolgsstatistiken unzureichend, denn sie stützten sich im Geist kameralistischen

3 BURDA, SEELE 2017; SVR 2017.

Denkens inputorientiert auf den erfolgreichen Mittelabruf und ließen dort, wo sie outputorientiert als Eingliederungs- oder Verbleibsquoten formuliert waren, die an sich zur Beurteilung erforderliche Erfassung von Netto-Wirkungen außen vor.

Dieser Auslöser und natürlich ursächlich der hohe Problemdruck durch die verfestigte Arbeitslosigkeit führten zu einer umfassenden Diskussion möglicher Reformen der Arbeitsmarktordnung im Rahmen der Hartz-Kommission. Intellektuell wurde diese Arbeit zweifellos durch den vorgelagerten wissenschaftlichen Diskurs vorbereitet, u.a. in den Arbeiten des Sachverständigenrates. Schließlich wurde die Verpflichtung zur zeitnahen und unabhängig nach dem aktuellen Stand der Wissenschaft durchgeführten Evaluation in die arbeitsmarktpolitischen Reformgesetze aufgenommen. Auf der Basis eines umfassenden und hinreichend praktikablen Reformdesigns wurden die ausgeschriebenen Evaluationspakete aufeinander abgestimmt.

Ein entscheidendes Element für den Erfolg dieser Bemühungen war der mit der gesetzlichen Verpflichtung einhergehende deutlich verbesserte Zugang zu den Prozessdaten der Bundesagentur für Arbeit. Zudem wurde der Wissenschaft die Möglichkeit eröffnet, diese aus verwaltungstechnischen und rechtlichen Gründen zwar akkurat erhobenen, aber informationsarmen Daten durch zusätzliche Erhebungen auf der Individual-ebene der Arbeitsmarktteilnehmer anzureichern. So konnte es gelingen, das ansonsten für nicht-experimentelle Ansätze der Evaluation häufig überwältigende Problem der unbeobachteten Heterogenität einzudämmen: Das übergreifende Prinzip jeder Erfolgsmessung ist schließlich „Vergleiche das Vergleichbare“.

Mittlerweile ist der Arbeitsmarkt – wohl nicht zuletzt aufgrund des Erfolgs dieser Reformen – in Deutschland etwas aus dem Blickfeld wirtschaftspolitischer Überlegungen gerutscht. Zum einen dürfte dies am gewaltigen Rückgang der Arbeitslosigkeit liegen, der im Gegensatz zu vielen anderen Volkswirtschaften selbst in der Großen Rezession nicht nennenswert unterbrochen wurde. Zum anderen ziehen mit der Digitalisierung und dem demographischen Wandel neue Probleme den Blick auf sich. Leider zeigt der aktuelle arbeitsmarktpolitische Diskurs, dass dadurch zentrale Erkenntnisse der Evaluationsstudien – etwa: Arbeitsbeschaffungsmaßnahmen schaden tendenziell den Geförderten sogar – offenbar vielfach in Vergessenheit geraten sind.

2. Wirtschaftspolitik: Ein informierter Prozess von Versuch und Irrtum

Wirtschaftspolitik setzt den Rahmen für das Handeln der Wirtschaftsakteure und greift zuweilen gezielt in dieses Geschehen ein. Die Lebenswirklichkeit, auf die sie Einfluss zu nehmen sucht, ist nicht weniger als ein Kaleidoskop aus vielfältigen Verbindungen zwischen diesen Akteuren, das stetig von einer schier unendlichen Zahl von Einzelentscheidungen verändert wird. Dabei folgen diese Entscheidungen vielfach Wünschen, Vorstellungen und Vorlieben, die nicht von der Politik zu beeinflussen sind, und sind häufig durch strategische Überlegungen geprägt. Das von der Wirtschaftspolitik beeinflusste Geschehen ist daher äußerst komplex und die Wirtschaftspolitik kann bei ihrer Einflussnahme nur begrenzte Wirkungen auslösen.

Darüber hinaus kann sie sich in vielen Fällen noch nicht einmal sicher sein, ob die von ihr ausgelösten Wirkungen in die gewünschte Richtung weisen und ob ihre Maßnahmen an anderer Stelle nicht unerwünschte Nebenwirkungen auslösen. Angesichts dieser unausweichlichen Unwägbarkeiten ist es unverzichtbar, die tatsächliche Wirkung von wirtschaftspolitischen Reformen und gezielten wirtschaftspolitischen Eingriffen sorgfältig zu evaluieren und auf Basis der dabei gewonnenen Erkenntnisse das politische Handeln anzupassen. Letztlich sollte es einer verantwortungsbewussten Wirtschaftspolitik darum gehen, nicht nur einen wohldurchdachten, sondern auch durch empirische Belege unterfütterten Entdeckungsprozess zu verfolgen.⁴

Der Beitrag der Ökonomik

Es ist der Kern der Ökonomik, menschliches Verhalten unter Knappheitsbedingungen zu erklären und vorherzusagen. Wirtschaftspolitische Eingriffe verändern die Rahmenbedingungen dieser Entscheidungen, indem sie Ressourcen verknappen oder zusätzliche Ressourcen – Informationen, finanzielle Mittel oder Zeit – bereitstellen. Die Wirtschaftswissenschaft ist heute besser denn je in der Lage, die durch diese Eingriffe ausgelösten Wirkungen zu analysieren. Sie kann sich dabei gleichermaßen auf eine eng mit dem Verständnis der Kausalität verbun-

4 Bei dieser Skizze einer evidenzbasierten Wirtschaftspolitik stütze ich mich nicht zuletzt auf SCHMIDT 2014.

dene tiefere konzeptionelle Durchdringung von Evaluationsproblemen und auf ein verbessertes statistisch-ökonometrisches Instrumentarium stützen.

Im Kern geht es bei Evaluationsstudien immer darum, die beobachtete Lebenswirklichkeit *mit* dem wirtschaftspolitischen Eingriff mit einer hypothetischen Situation *ohne* diesen Eingriff zu kontrastieren: Die Differenz zwischen beiden misst die kausale Wirkung des Eingriffs. Die Möglichkeiten dazu sind besser denn je, denn die ökonometrischen Methoden wurden schritthaltend mit der steigenden Rechenkapazität und dem besseren Zugang zu umfangreichen Daten ausgebaut. Dieser Fortschritt ist nicht aufzuhalten, der nächste Entwicklungsschritt wird die Analyse immer größerer Datenmengen mithilfe der Methoden der künstlichen Intelligenz einbeziehen. Doch viele Informationen einbeziehen zu können, ist keineswegs hinreichend.

Denn der entscheidende Fortschritt der Evaluationsforschung liegt im tieferen konzeptionellen Verständnis dieser Aufgabe: Es gibt letztlich keine sicheren Aussagen, die sich durch den einfachen Vergleich der beobachteten Lebenswirklichkeit in zwei unterschiedlichen Zeiträumen, Regionen etc. — von statistischer Restunsicherheit abgesehen — mit Gewissheit ableiten ließen. Vielmehr muss die als Kontrast herangezogene hypothetische Vergleichssituation immer bewusst konstruiert werden. Um tatsächlich aus diesem Kontrast die ursächliche Wirkung eines Eingriffs ablesen zu können, muss die Vergleichssituation in allen relevanten Belangen — bis auf die zu bewertende Maßnahme — mit der beobachteten Lebenswirklichkeit übereinstimmen.

Aus dieser grundlegenden Aufgabe kann sich niemand befreien. Sie schlicht zu ignorieren und aus der Intuition heraus einen impliziten Vergleichspunkt zu wählen, bspw. die Ausgangssituation vor der Maßnahme, ist zwar häufig intuitiv naheliegend, aber kein vernünftiger Lösungsvorschlag. Denn diese intuitive Wahl dürfte nur selten zufällig die Anforderung der Vergleichbarkeit erfüllen. Im schlimmsten Falle führt die implizite Wahl einer Vergleichssituation sogar zu völlig falschen Schlussfolgerungen. Die Gretchenfrage an eine jede Evaluationsstudie ist demnach, ob das gewählte Studiendesign tatsächlich diese zentrale Anforderung erfüllt, die gesuchte Vergleichssituation in überzeugender Weise zu konstruieren.

Die Natur- und Technikwissenschaften haben darauf eine hervor-

ragende Antwort entwickelt: Sie stützen sich auf kontrollierte Zufallsexperimente. Durch die zufallsgestützte Zuweisung unterschiedlicher „Behandlungsregimes“ verlieren überlagernde Einflüsse an Bedeutung, und der gesuchte Kontrast mit und ohne Maßnahme tritt in den Vordergrund. Daher nutzen Ökonomen vermehrt die Möglichkeit, ebenfalls kontrollierte Experimente durchzuführen. Noch häufiger dient ihnen allerdings das nicht durchführbare Experiment, das in dem von ihnen untersuchten Kontext das ideale Studiendesign darstellen würde, als gedanklicher Anker des von ihnen in der praktischen Arbeit eingesetzten nicht-experimentellen Studiendesigns.

Umsetzungshemmnisse in der wirtschaftspolitischen Praxis

Denn oft lassen sich diese Einsichten in die Natur eines idealen Studiendesigns nicht in die praktische Analyse übersetzen. Vor allem erfordert ein experimentelles Studiendesign, dass die Wissenschaft bereits lange vor dem Zeitpunkt eingebunden werden muss, an dem die Umsetzung der Maßnahme beginnt. In der wirtschaftspolitischen Praxis suchen Politik, Verwaltung und Öffentlichkeit aber häufig erst *nach* dem Beginn der Umsetzung wissenschaftlichen Rat. Häufig genug geschieht dies nicht aus reiner Nachlässigkeit, sondern weil sie eine experimentelle Analyse aus grundlegenden Erwägungen heraus ablehnen. Jedenfalls lassen sich überlagernde Einflüsse dann nicht mehr per Zufallsauswahl aus dem Bild nehmen.

Es ist daher die typische Aufgabe der angewandten Evaluationsforschung, mit nicht-experimentellen Ansätzen dasjenige Experiment so gut wie möglich anzunähern, das man im Ideal gern von vornherein durchgeführt hätte. Die moderne Mikro-Ökonometrie bietet zu diesem Zweck ein sehr leistungsfähiges Instrumentarium. Nicht zuletzt ist sie in der Lage, vielfältige andere Einflussfaktoren so zu berücksichtigen, dass die Wirkung der Maßnahme weitgehend isoliert abgeschätzt werden kann. Von besonderer Bedeutung für die praktische Arbeit sind dabei die vertrauensvolle Zusammenarbeit der Wissenschaft mit Politik und Verwaltung und der damit verbundene Zugang zu umfassenden Individualdaten.

Allerdings gilt es für die Wissenschaft immer, die Grenzen des gewählten Studiendesigns zu erkennen und dementsprechend die ermittelten Ergebnisse mit großer Bescheidenheit vorzutragen. Es handelt sich schließlich um empirische Belege, die nur unter erheblichen Vorbedin-

gungen gewonnen werden können: Diese „Identifikationsannahmen“ sind die entscheidende Brücke von der Berechnung von bloßen Korrelationen hin zu Aussagen im Hinblick auf kausale Zusammenhänge. Bei einer evidenzbasierten Politik(-Beratung) kommt es daher nicht allein darauf an, Entscheidungen auf empirische Belege zu stützen. Entscheidend ist vielmehr, diese Identifikationsannahmen intelligent einzusetzen und transparent offenzulegen.⁵

Eine Einsicht, die für Ökonomen Allgemeingut ist, erweist sich hier erneut als zutreffend: Das Leben ist voller Zielkonflikte. In der Praxis hat sich als Erfahrungswissen herausgeschält, dass es typischerweise besonders schwer ist, jene wirtschaftspolitischen Eingriffe im Hinblick auf ihre tatsächliche Wirkung zu bewerten, die flächendeckend wirksam werden, die vielfältige Umsetzungselemente enthalten und die von erheblichen anderen Einflüssen überlagert werden.

- Werden Maßnahmen flächendeckend umgesetzt, so ist es tendenziell umso schwerer, im Untersuchungskontext ein hinreichend großes Reservoir an Vergleichsbeobachtungen zu finden. Im schlimmsten Falle entzieht sich die Maßnahme somit gänzlich einer Evaluation.
- Bestehen Maßnahmen aus einer Fülle von Elementen, die mal mehr und mal weniger zum Einsatz kommen, dann wird es schwerfallen, die Wirkung der einzelnen Elemente zu isolieren. Im schlimmsten Falle lassen sich daher weder produktive noch kontraproduktive Elemente herausfiltern.
- Schließlich ist es eine besondere Herausforderung, die Wirkungen von Maßnahmen empirisch zu isolieren, wenn sie in einer Situation umgesetzt werden, die sich ohnehin in einem allgemeinen Umbruch befindet. Im schlimmsten Falle wird der Strukturbruch als Wirkung der Maßnahme missverstanden.

Doch genau diese Reformpakete mit großem Geltungsbereich, die umfassende Veränderungen auf vielen Ebenen mit sich bringen, bieten aus

5 Diese Einsichten haben mittlerweile Einzug in einschlägige Lehrbücher der empirischen Wirtschaftsforschung gefunden; exemplarisch sei hier auf BAUER ET AL. 2009 verwiesen.

wirtschaftspolitischer Sicht oft die spannendsten Evaluationsaufgaben. Sie kommen noch dazu meist erst dann zum Einsatz, wenn die Zeiten turbulent und die Herausforderungen besonders groß sind. Gerade dann wird es besonders relevant, verlässlich abschätzen zu können, welche wirtschaftspolitischen Eingriffe funktionieren und welche nicht. Dies alles traf auf das umfassende Reformpaket zu, das zu Beginn dieses Jahrhunderts darauf abzielte, die deutsche Arbeitsmarktordnung und die Funktionsfähigkeit der arbeitsmarktpolitischen Instrumente und Institutionen grundlegend zu verändern.

Die Wissenschaft darf einer solchen Aufgabe nicht ausweichen, indem sie sich auf wirtschaftspolitisch weniger spannende, analytisch jedoch leichter zu bewältigende Evaluationsaufgaben konzentriert. Sie ist vielmehr aufgerufen, ausgehend von ihrem konzeptionellen Verständnis nach Lösungsansätzen zu suchen, die zwar praktikabel und verständlich zu erklären sind, aber dennoch den wissenschaftlichen Anspruch der Wahrhaftigkeit und Sorgfalt nicht verletzen. Wie dies gelingen kann, zeigt exemplarisch eine Studie,⁶ die im Auftrag des Bundesministeriums für Wirtschaft und Arbeit (BMWA) ein umfassendes Konzept für die Evaluation der Hartz-Gesetze zur Arbeitsmarktpolitik erarbeitete.⁷

3. Die Hartz-Reformen: Zielsetzungen und konzeptionelle Probleme

Die Hartz-Reformen stellen für Deutschland das größte arbeitsmarktpolitische Reformpaket der vergangenen Jahrzehnte dar. Kaum ein Aspekt der Arbeitsmarktordnung blieb unberührt und der arbeitsmarktpolitische Instrumentenkasten wurde erheblich überholt. Die Reformen zielten darauf ab, eine raschere und nachhaltigere Vermittlung in Arbeit zu bewirken, neue Beschäftigungsfelder zu schaffen und mehr Brücken in Beschäftigung zu bauen. Ein Kernbaustein war die Idee des „Fördern und Fordern“: Optimierte staatliche Unterstützung für arbeitssuchende Menschen sollte mit der Stärkung ihrer Eigenverantwortlichkeit einher-

6 FERTIG ET AL. 2004.

7 Diese Studie wurde im August 2003 vom damaligen Bundesministerium für Wirtschaft und Arbeit (BMWA) in Auftrag gegeben. Das BMWA gab zeitgleich eine weitere Studie in Auftrag, welche die gleiche Zielsetzung verfolgte (HAGEN, SPERMANN 2004). Aufbauend auf den Empfehlungen dieser beiden Vorstudien wurden in den Jahren danach die eigentlichen Evaluationsstudien vergeben. Eine Übersicht über die Ergebnisse dieser Studien geben JACOBI, KLUVE 2007.

gehen. Der übergreifende Anspruch des Reformpakets war letztlich die Verringerung der Arbeitslosigkeit. Die Frage nach seiner Wirksamkeit ist somit eine „große“ Frage, die im Gegensatz zur Wirkung isolierter und inkrementeller Reformen nur schwer zu beantworten ist.⁸

Elemente des Reformpakets

Die Gesamtheit dieser Reformen ließ sich in zwei große Gruppen unterteilen. Die erste Gruppe umfasste die Veränderung oder Einführung konkreter Instrumente der Arbeitsmarktpolitik, die auf einzelne Arbeitnehmerinnen und Arbeitnehmer oder einzelne Unternehmen abzielen. Dazu zählten Instrumente wie die Eingliederungszuschüsse und die Ich-AG, die Förderung der beruflichen Weiterbildung, Eingliederungsmaßnahmen durch Träger sowie Änderungen im Arbeitnehmerüberlassungsgesetz und der Regelungen über befristete Arbeitsverhältnisse. Ebenso zählte dazu die Neuregelung geringfügiger Beschäftigungsverhältnisse („Mini-“ und „Midi-Jobs“), die Einführung von Personal-Service-Agenturen sowie von Vermittlungsgutscheinen.

Die zweite Gruppe an Reformen veränderte die Rahmenbedingungen der Arbeitsmarktpolitik. Dazu zählten zum einen Veränderungen ihrer Zielsetzung, Administration und Steuerung; so wurden bspw. Arbeitslosengeld und Unterhaltsgeld zusammengelegt, die Bemessung des Arbeitslosengeldes vereinfacht und Arbeitsbeschaffungs- und Strukturanpassungsmaßnahmen zusammengeführt. Darüber hinaus führten die Reformen ein ergebnisorientiertes Steuerungskonzept ein und etablierten Job-Center als einheitliche Anlaufstellen für alle Arbeitsuchenden.

Zum anderen wurde das Leistungsrecht überholt. So wurde die Bezugsdauer des Arbeitslosengelds verkürzt, eine einheitliche Anwartschaftszeit eingeführt und die Sperrzeitregelung neu formuliert, die Meldepflicht verändert und die Zumutbarkeitsvorschriften verschärft. Darüber hinaus wurden Arbeitslosenhilfe und Sozialhilfe zur „Grundsicherung für Arbeitsuchende“ zusammengeführt. Die Veränderungen in den Rahmenbedingungen zielten letztlich insgesamt darauf ab, die auf individuelle Akteure ausgerichteten Instrumente der Arbeitsmarktpolitik wirksam zu unterstützen.

8 Die Abschnitte 3 und 4 dieses Beitrags sind eng an die Diskussion in FERTIG ET AL. 2004 angelehnt.

Das Ineinandergreifen von Reformkomponenten

Diese einzelnen Reformen ließen sich nicht ohne Weiteres getrennt voneinander evaluieren, da sie inhaltlich und in ihrer Zielsetzung ineinandergriffen und daher als ein Ganzes betrachtet und evaluiert werden mussten. Nichtsdestoweniger waren die Evaluationsbemühungen ebenfalls darauf auszurichten, so gut wie möglich den Erfolg einzelner Eingriffe zu isolieren und zu quantifizieren. Denn nur dann lässt sich wirksam verhindern, dass sich die Wirkungen mehrerer Reformelemente in der wirtschaftspolitischen Einschätzung überlagern. In der wirtschaftspolitischen Praxis ist die Gleichzeitigkeit verschiedener Eingriffe zwar eher die Regel als die Ausnahme. Bei den Hartz-Reformen wog dieses Problem allerdings besonders schwer.

Bei den Reformen der Zielsetzung, Administration und Steuerung der Arbeitsmarktpolitik war insbesondere das neue, ergebnisorientierte Steuerungskonzept der Arbeitsmarktpolitik von Bedeutung. Anstelle der bisherigen „inputorientierten“ Steuerung sollte es – zusammen mit Vereinfachungen im Verwaltungsablauf – der Bundesagentur für Arbeit ermöglichen, sich auf ihre Kernaufgaben zu konzentrieren: die Vermittlung in Arbeit und die Gewährung vermittlungsfördernder Leistungen.

Zudem sollte den Mitarbeiterinnen und Mitarbeitern der Arbeitsagenturen neuer Freiraum für eine intensivere Betreuung ihrer Schützlinge eröffnet werden. Davon versprach man sich nicht nur eine höhere Betreuungsintensität und darauf aufbauend eine schnellere Vermittlung von Arbeitslosen in Arbeit, sondern auch die passgenauere Bereitstellung unterstützender Leistungen und darauf aufbauend mehr Erfolg der auf individuelle Akteure ausgerichteten Instrumente der Arbeitsmarktpolitik.

Diese allgemeine Veränderung der Rahmenbedingungen überlagerte natürlich jeden Vergleich der Wiedereingliederungserfolge einzelner Instrumente *vor* den Hartz-Reformen mit denen *nach* ihrer Umsetzung. Darüber hinaus war zu erwarten, dass die Umsetzung des neuen Steuerungskonzepts über die Arbeitsämter hinweg variieren und somit zu Unterschieden in der tatsächlichen Betreuungsintensität führen würde. Die konkrete Umsetzung des neuen Steuerungskonzepts vor Ort musste daher bei der Bestimmung des Erfolgs einzelner Instrumente der Arbeitsmarktpolitik berücksichtigt werden, um ihre ursächliche Wirkung isoliert von der Veränderung der Rahmenbedingungen abschätzen zu können.

Als zentrales Element der Evaluationsstudien war somit die Erhebung von Indikatoren für die Qualität der Dienstleistungserbringung in den einzelnen Arbeitsagenturen vorzusehen. Mögliche Indikatoren umfassen beispielsweise für jede Arbeitsagentur die Anzahl der Vermittlerinnen und Vermittler je arbeitslos Gemeldetem, die Anzahl der Online-Vermittlungsplätze je arbeitslos Gemeldetem oder die durchschnittliche Dauer des Gesprächs zwischen Arbeitslosen und ihren Vermittlern vor Ort.

Veränderungen im Leistungsrecht können ebenfalls Rückwirkungen auf den Erfolg konkreter Instrumente der Arbeitsmarktpolitik haben. So könnten die verkürzte Arbeitslosengeldbezugsdauer oder die verschärfte Zumutbarkeitsregelung dazu geführt haben, dass Arbeitslose sich stärker um eine Beschäftigung bemühen oder eine Weiterbildungsmaßnahme mit höherer Motivation absolvieren. Daher waren die Erwerbsbiographie von Arbeitslosen sowie deren familiäre Situation detailliert abzubilden. Denn Ansprüche auf Entgeltersatzleistungen und die Dringlichkeit einer raschen Wiederbeschäftigung richteten sich vor allem an diesen Rahmenbedingungen aus.

Allgemeingültige Instrumente

Wir erinnern uns: Die ursächliche Wirkung eines wirtschaftspolitischen Eingriffs lässt sich nur durch den Kontrast zweier Situationen ermitteln, die sich bis auf einen einzigen Umstand sehr ähneln: In der einen Situation wird die Intervention durchgeführt und in der anderen nicht. Das übergreifende Prinzip ist „Vergleiche das Vergleichbare“. Bei den Hartz-Reformen fiel die Konstruktion dieses Kontrasts jedoch für solche neu eingeführten arbeitsmarktpolitischen Instrumente besonders schwer, die allen Akteuren den gleichen Zugang gewährten. Dies galt insbesondere für den Rechtsanspruch auf den Existenzgründerzuschuss und das Überbrückungsgeld, den Zugang zu Mini- und Midi-Jobs sowie das reformierte System der Grundsicherung.

Das Reservoir potenzieller Vergleichsbeobachtungen, die von diesen allgemeingültigen Instrumenten ausgeschlossen gewesen wären, war somit zumindest dem ersten Eindruck nach leer. Zum gleichen Zeitpunkt und in der gleichen Gruppe von Arbeitnehmern und Unternehmen ließen sie sich zumindest nicht direkt finden. Dem Prinzip „Vergleiche das Vergleichbare“ zu genügen, fiel somit besonders schwer. Es galt

also nach alternativen Vergleichsgruppen zu fahnden, die als Ersatz für das Reservoir an von der Maßnahme ausgeschlossenen Akteure dienen konnten, bspw. durch einen Vergleich unterschiedlicher Zeitpunkte. Es ist die ureigene Aufgabe der Wissenschaft, auch in dergestalt komplexen Situationen Lösungskonzepte zu erarbeiten und deren jeweils gewählte Vergleichbarkeitsannahme überzeugend zu begründen.

4. Die Hartz-Reformen: umfassendes Evaluationskonzept

Vor diesem Hintergrund entwickelte ein Konsortium aus ISG⁹ und RWI¹⁰ im Jahr 2003 ein in sich geschlossenes Konzept zur mittel- und längerfristigen Bewertung der Auswirkungen des Reformpakets, der Effektivität und Effizienz einzelner Maßnahmen sowie zur Analyse der hinter den ermittelten Wirkungen stehenden Gründe.¹¹ Eng verwoben mit den konzeptionellen Problemen der umfassenden und ineinandergreifenden Natur der Maßnahmen waren dabei zwei praktische Probleme zu bedenken: Die Bewertung der Maßnahmen musste das gleichzeitig eingebrachte neue Steuerungskonzept berücksichtigen und den zeitnahen Zugang zu den für die gewählten empirischen Ansätze relevanten Daten sicherstellen.

Praktische Probleme der Evaluation

An diesem Beispiel zeigte sich lebhaft, welche erheblichen praktischen Probleme die Evaluation eines umfassenden Reformpakets aufwirft, insbesondere dann, wenn sie zeitnah durchgeführt werden und die Politik zu einer raschen Nachsteuerung befähigen soll. So waren zum Zeitpunkt der Evaluation die konkrete Ausgestaltung des neuen Steuerungskonzepts sowie der Stand der Umsetzung in der Bundesagentur für Arbeit noch weitgehend unklar. Dies betraf insbesondere die über die Kunden der Arbeitsagenturen gesammelten Daten und ihre mögliche Bereitstellung für den Zweck der Evaluation sowie den möglichen Katalog an Indikatoren für die Qualität der Dienstleistungserbringung vor Ort.

Von entscheidender Bedeutung für die Qualität jeder Evaluationsstu-

9 ISG – Institut für Sozialforschung und Gesellschaftspolitik Köln.

10 RWI – Leibniz-Institut für Wirtschaftsforschung, damals Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

11 FERTIG ET AL. 2004.

die ist die zeitnahe Verfügbarkeit informativer und wenig revisionsanfälliger Daten. Den Evaluierenden einen möglichst freimütigen Zugang zu allen Verwaltungsdaten auf der Individualebene zu gewähren, ist ein guter erster Schritt, reicht aber für sich genommen nicht aus. Denn die Daten der Arbeitsverwaltung stehen häufig nur mit einem gewissen zeitlichen Abstand zum Geschehen in hinreichender Qualität zur Verfügung. Diese Verzögerung lässt sich nicht beliebig verringern. Daher ist es typischerweise notwendig, das verfügbare Datenmaterial um eigene Erhebungen zu ergänzen, nicht zuletzt im Hinblick auf Details zu den durchgeführten Maßnahmen selbst.

Bei den Hartz-Reformen hat es sich als hilfreich erwiesen, dass auf der Ebene individueller Arbeitnehmerinnen und Arbeitnehmer viele der für die Evaluation wichtigen sozioökonomischen Eigenschaften in den Verwaltungsdaten vorlagen. Darüber hinaus ließen sich aus ihnen grundsätzlich die aus konzeptioneller Sicht zentralen Informationen zur Erwerbsbiographie der Arbeitnehmerinnen und Arbeitnehmer extrahieren. Eigene Erhebungen konnten daher genutzt werden, um diesen Informationsstand sinnvoll zu ergänzen, bspw. um Informationen zum aktuellen Arbeitsmarktstatus des gegebenenfalls vorhandenen Lebenspartners oder zu den Kindern.

Sorgfältige Datenerhebungen beinhalten vorbereitende Arbeiten, die Versendung der Fragebögen und entsprechende Nachfassaktionen. Für die eigentliche Auswertung der Daten wird ebenfalls eine gewisse Zeit benötigt. Von der geplanten Vorlage des Evaluationsberichts konnte man daher leicht auf den Zeitpunkt zurückschließen, an dem mit der ergänzenden Datenerhebung begonnen werden musste. Da einige Reformkomponenten typischerweise erst mit einer gewissen zeitlichen Verzögerung in Kraft treten können, musste das Evaluationsdesign zulassen, dass die Evaluationsbemühungen im Zweifelsfalle zeitlich ausgedehnt werden und im Zeitablauf neu ermittelte Resultate nahtlos in das Gesamtbild eingefügt werden konnten.

Als Ergänzung der in diesem Rahmen gewonnenen Erkenntnisse über die Wirkung einzelner arbeitsmarktpolitischer Instrumente sowie ihrer Veränderungen im Zuge der Hartz-Reformen auf mikroökonomischer Ebene schlugen Fertig et al. (2004) vor, die Arbeitsmarktpolitik als Ganzes auf Ebene der Arbeitsagenturbezirke zu analysieren. Dafür sprach nicht zuletzt, dass diese Untersuchung im Prinzip auf alle bis dahin verfügba-

ren Daten aus den Eingliederungsbilanzen der Arbeitsämter und den Eckdaten des Arbeitsmarktes zurückgreifen konnte. Zudem ergab sich damit ein natürlicher Ansatzpunkt, um die zu diesem Zeitpunkt noch nicht systematisch erfassten Indikatoren für die Qualität der Dienstleistungserbringung zu ergänzen.

Ein weiteres praktisches Problem bei der Evaluation von Arbeitsmarktmaßnahmen ergibt sich aus der gleichzeitigen Verfügbarkeit anderer Unterstützungsleistungen. Dies galt bspw. für Förderangebote des Europäischen Sozialfonds (ESF): Einzelne Bundesländer hatten die Förderung von Arbeitsbeschaffungsmaßnahmen durch aus dem ESF finanzierte Qualifizierungsmaßnahmen ergänzt, ohne dass diese „Doppelförderung“ in den Verwaltungsdaten der Bundesagentur für Arbeit explizit erfasst wurde. Um zu vermeiden, dass die Schätzungen zur Wirkung der arbeitsmarktpolitischen Instrumente durch eine solche Doppelförderung verzerrt wurden, waren auch diese Informationen im Rahmen von zusätzlichen Erhebungen gezielt zu erfassen.

Leitlinien des Evaluationskonzepts

Vor diesem Hintergrund hatten ISG und RWI ein pragmatisches Evaluationskonzept vorgeschlagen, das sich erfolgreich an den folgenden fünf Leitlinien orientierte:¹²

► ***Umsetzbarkeit.***

Das Konzept musste angesichts des begrenzten Datenmaterials und des kurzen Zeitfensters, das für die Durchführung der Evaluationsstudien zur Verfügung stand, ohne größere Probleme umzusetzen sein.

► ***Fokussierung.***

Das Konzept musste sich angesichts der Vielzahl der (Teil-)Ziele der Arbeitsmarktpolitik und der im Rahmen der Hartz-Reformen durchgeführten Veränderungen auf einen begrenzten Katalog an Kernzielen konzentrieren.

¹² Hier wird allerdings die dort ebenfalls genannte Anforderung der Interdisziplinarität zum Zwecke der transparenten Darstellung ausgeblendet.

► *Einheitlichkeit.*

Das Konzept musste trotz der Heterogenität des Reformpakets einen einheitlichen Analyserahmen anbieten, der (i) die Effektivität und (ii) die Effizienz der Maßnahmen und (iii) die Ursachen hinter den Ergebnissen anspricht.

► *Robustheit.*

Das Konzept musste Methoden vorsehen, die mit dem vorhandenen Datenmaterial und innerhalb des engen Zeitrahmens Ergebnisse erzeugen konnten, die hinsichtlich einer begrenzten Variation der konkreten Ansätze robust sind.

► *Erweiterbarkeit.*

Das Konzept musste so angelegt sein, dass es ohne größere Anpassungen um später beschlossene oder veränderte Maßnahmen erweitert werden und im Prinzip als Blaupause für künftige Reformpakete dienen konnte.

Kernelemente der praktischen Umsetzung des vorgeschlagenen Konzepts waren (i) die Erfassung und Auswertung zentraler statistischer Begleitdaten („Monitoring“), (ii) eine qualitative Analyse der Verbesserung der Dienstleistungen der Arbeitsverwaltung und (iii) quantitative mikro- und makroökonomische Analysen einzelner Aspekte der Arbeitsmarktpolitik und der Arbeitsmarktpolitik insgesamt. Damit war der Nachweis erbracht, dass es selbst bei hoher zeitlicher Dringlichkeit des politischen Handelns möglich ist, zeitnah Evaluationsergebnisse für komplexe Reformpakete vorzulegen, die sowohl hohen wissenschaftlichen Ansprüchen genügen als auch robust genug sind, um die informierte Anpassung der Reformelemente zu erlauben.

5. Ausblick

Nach wie vor sehen die für Wirtschaftspolitik verantwortlichen Akteure in wissenschaftlichen Analysen zu selten die Chance, die Wirksamkeit ihres Handelns zu verbessern und ihm noch dazu durch den wissenschaftlich gestützten Erfolgsnachweis eine zusätzliche Legitimation zu verleihen. Die Sorge, dass eine Maßnahme oder eine Reform nicht wie gewünscht wirkt, müsste doch aufgrund der Anstrengungen bei deren

Design ohnehin begrenzt sein. Andererseits sind Hemmungen der Programmverantwortlichen nachvollziehbar, ihre mit viel Herzblut implementierten Programme von Dritten möglicherweise negativ bewertet zu sehen, zumal sie selbst letztlich wenig Einfluss auf die tatsächliche Wirkung nehmen können.

Doch hier geht es um viel, denn die von diesen Reformen angesprochenen Problemlagen sind häufig von erheblicher wirtschaftspolitischer und gesellschaftlicher Bedeutung. Nicht zuletzt berühren sie das Lebensglück vieler einzelner Menschen in erheblichem Maße; wirksame Reformen zu erkennen und unwirksame zu enttarnen, ist daher ein wichtiges Element verantwortungsbewussten und klugen wirtschaftspolitischen Handelns. Die Wissenschaft hat die Aufgabe, in Politik, Verwaltung und Öffentlichkeit stetig weiter Überzeugungsarbeit für die völlig zutreffende Einsicht zu leisten, dass der wirtschaftspolitische Entdeckungsprozess durch Evaluationsanstrengungen erfolgreicher zu gestalten ist.

Um dies zu schaffen, müssen beide Seiten dieses Dialogs ihre Anstrengungen zum gegenseitigen Verständnis verstärken. Je höher die methodische Kompetenz auf der Nachfrageseite entwickelt ist, umso eher operieren Politik und Verwaltung einerseits und Wissenschaft andererseits auf Augenhöhe. Insbesondere besteht die Hoffnung, dass ein besseres Verständnis der Nachfrager über Möglichkeiten und Grenzen der Evaluation es umso wahrscheinlicher macht, dass sie die Voraussetzungen für eine Evaluation von wirtschaftspolitischen Maßnahmen bereits bei deren Design und Implementierung mit bedenken. Das wichtigste Element dieser Voraussetzungen ist zweifellos der praktikable Zugang zu reichhaltigem, mit der Aufgabenstellung verknüpftem Datenmaterial.

Wichtig auf der Seite von Politik und Verwaltung sind somit (i) die Offenheit, die eigene Arbeit von Außenstehenden kritisch bewerten zu lassen, (ii) die Bereitschaft, freimütigen Datenzugang – also zum hoheitlichen Besitzstand – zu gewähren, und (iii) die Souveränität, die Komplexität der Angelegenheit anzuerkennen: Wissenschaft hat nie ganz sichere Antworten, die Wirkungen der Programme sind heterogen, Erfolg kann immer nur eine Verschiebung von Erfolgswahrscheinlichkeiten sein. Um ein fundiertes eigenes Urteil kommt die Politik somit nicht herum.

Wichtig auf der Seite der Wissenschaft sind somit (i) die Bereitschaft, auf der Suche nach konkreten Antworten pragmatische Annäherungen an das ideale Studiendesign zu suchen, ohne die wissenschaftliche Basis

zu kompromittieren, (ii) die Souveränität, die Grenzen der eigenen Erkenntnis freimütig zu bekennen, selbst wenn im wirtschaftspolitischen Diskurs andere Akteure bereit sind, die wissenschaftliche Basis zu verlassen, und (iii) das ernsthafte Bemühen, komplexe Zusammenhänge so einfach und transparent wie möglich darzustellen.

Literatur

Bauer, Th.K., Fertig, M., Schmidt, Ch.M. (2009). „Empirische Wirtschaftsforschung: Eine Einführung“. Berlin: Springer.

Burda, M.C., Seele, S. (2017). „Das deutsche Arbeitsmarktwunder: Eine Bilanz“. Perspektiven der Wirtschaftspolitik 18:3, 179-204.

Deutsche Bundesbank (2017). „Finanzstabilitätsbericht 2017“. Frankfurt am Main.

Fertig, M., Kluve, J., Schmidt, Ch.M., Apel, H., Friedrich, W., Hägele H. (2004). „Die Hartz-Gesetze zur Arbeitsmarktpolitik: Ein umfassendes Evaluationskonzept“. RWI-Schriften 74. Berlin: Duncker & Humblot.

FSB – Financial Stability Board (2017). „Framework for Post-Implementation Evaluation of the Effects of the G20 Financial Regulatory Reforms“. Basel.

Hagen, T., Spermann, A. (2004). „Hartz-Gesetze: Methodische Ansätze zu einer Evaluierung“. ZEW-Wirtschaftsanalysen 74. Baden-Baden: Nomos.

Jacobi, L., Kluve, J. (2007). „Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany“. Zeitschrift für ArbeitsmarktForschung 40:1, 45-64.

Schmidt, Ch.M. (2014). „Wirkungstreffer erzielen – Die Rolle der evidenzbasierten Politikberatung in einer aufgeklärten Gesellschaft“. Perspektiven der Wirtschaftspolitik 15:3, 219-233.

SVR – Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2017). „Für eine zukunftsorientierte Wirtschaftspolitik“. Jahresgutachten. Kapitel 8: Arbeitsmarkt: Fachkräftesicherung im digitalen Wandel. Wiesbaden, 360-406.

Evidenzbasierte Beratung im Politikfeld Arbeitsmarkt

Von ULRICH WALWEI

Korrespondenzadresse:

Institut für Arbeitsmarkt- und Berufsforschung (IAB) der Bundesagentur
für Arbeit (BA)

Regensburger Str. 104

90478 Nürnberg

Ulrich.Walwei@iab.de

1. Einleitung

Der Arbeitsmarkt gilt gemeinhin als gutes Beispiel für die Herausbildung einer evidenzbasierten Politikberatung.¹ Zu den Wirkungen arbeitsmarktpolitischer Interventionen liegt eine Vielzahl von Forschungsbefunden in beachtlicher Qualität vor. Damit steht den Entscheidungsträgern in Politik (z.B. in den für Arbeit zuständigen Bundes- und Landesministerien) und Praxis (z.B. in den Dienststellen der Bundesagentur für Arbeit) ein umfassendes Wissen zur Funktionsweise des Arbeitsmarktes und den Effekten von Maßnahmen zur Verfügung.

Die Existenz einer qualitativ hochwertigen Politikberatung, die sich auf wissenschaftliche Evidenz stützt, ist nicht selbstverständlich. Zwei Bedingungen sind essentiell: Zum einen muss die Wissenschaft fähig und in der Lage sein, kontinuierlich exzellente und gleichermaßen relevante Ergebnisse zu produzieren. Voraussetzungen dafür sind passende Rahmenbedingungen (u.a. zugängliche Daten und gut ausgebildete Wissenschaftlerinnen und Wissenschaftler) sowie wirksame Anreize (z.B. durch eine gezielte Forschungsförderung). Zum anderen müssen die Ergebnisse für die Gestaltung politischer und operativer Eingriffe verwertbar sein sowie Politik und Praxis willens und in der Lage sein, diese zu rezipieren und bei Umsetzungsentscheidungen zu berücksichtigen.

Der nachfolgende Beitrag beschäftigt sich mit drei Fragen. Worin besteht allgemein der Anspruch einer evidenzbasierten Politikberatung und wo liegen deren spezifische Ansatzpunkte? Worauf lässt sich die

1 ZIMMERMANN 2011.

hohe Intensität evidenzbasierter Politikberatung im Politikfeld Arbeitsmarkt zurückführen? Welche Lehren ergeben sich daraus für andere Politikfelder, die mit Blick auf wissenschaftliche Begleitung noch Nachholbedarf haben?

2. Anspruch und Ansatzpunkte evidenzbasierter Politikberatung

Um Ansprüche an eine evidenzbasierte Politikberatung zu formulieren, sind zunächst die beiden Akteursgruppen ins Blickfeld zu nehmen. Dabei geht es einerseits um die Adressaten (z.B. Ministerien oder deren nachgelagerte, zumeist ausführende Einrichtungen) und andererseits um die wissenschaftlichen Produzenten der evidenzbasierten Politikberatung. Im Vordergrund stehen deren jeweilige Interessenssphären und Rahmenbedingungen.

Für den Adressaten Politik besteht die Herausforderung darin, dass die wesentlichen Handlungsfelder vielfach strategischer und damit langfristiger Natur sind. Um diese konsequent zu bearbeiten, bedarf es nicht selten einer beträchtlichen, für die Wahlbevölkerung unter Umständen unpopulären Umsteuerung. Beispiele für solche Handlungsfelder sind der Klimawandel, die Folgen der demografischen Entwicklung oder auch die fortschreitende Digitalisierung. Gleichzeitig verfolgen Politikerinnen und Politiker – wenn die Annahmen der ökonomischen Theorie der Politik zugrunde gelegt werden – vorrangig das Ziel, politische Macht zu erlangen und diese zu festigen.² Die nächste Wahl ist dabei zumeist die handlungsleitende Maxime. Hieraus entstehen Konflikte, denn das Ringen um die Wählergunst legt es nicht unbedingt nahe, zuallererst die drängenden, oft schwierig zu lösenden Langfristprobleme anzugehen. Von daher gibt es die Tendenz in der Politik, kurzfristig orientierten und opportunen Problemlösungen den Vorzug zu geben. Um Wählerstimmen zu maximieren, sind sie bei Interessenskonflikten deshalb zuallererst um einen Interessensausgleich bemüht. Hierzu gehört, Kompromisse zu gestalten, Meinungen zu kanalisieren und mehrheitsfähig zu machen.

Auch auf Seiten der Forschung finden sich spezifische Interessenslagen. Wissenschaftlerinnen und Wissenschaftler streben in erster Linie nach Anerkennung in der Scientific Community. Dabei geht es nicht

2 DOWNS 1957; BUCHANAN, TULLOCK 1962; DEHLING, SCHUBERT 2011.

allein um Prestige und Selbstbestätigung, sondern auch um Karriere und Entlohnung.³ Entscheidend für das individuelle Vorankommen in der Wissenschaft sind der Nachweis der in der jeweiligen Disziplin anerkannten Publikationen und die Einbindung in relevante Forschungsnetzwerke. Forschung ist deshalb nicht per se gut auf eine wissenschaftsorientierte Politikberatung vorbereitet. Insbesondere sichern die bestehenden Anreizstrukturen nicht die für die wissenschaftliche Politikberatung notwendige Relevanz. Gemeint ist damit die Nähe zur Praxis und zum politischen Tagesgeschäft sowie zu den Fragestellungen, die bei einer konkreten Umsetzung von Maßnahmen, z.B. in Form von Gesetzen, zu bedenken sind. Dazu kommt, dass durchaus Evidenz vorliegen kann und prinzipiell auch abrufbar wäre, aber eben nicht zu den Fragen, die gerade seitens der Entscheidungsträger gestellt werden. Der Nutzen vorliegender Forschungsbefunde kann zudem dadurch beeinträchtigt sein, dass diese nur für einen weit zurückliegenden Zeitraum vorhanden sind und deshalb für den aktuellen Rand ggf. nur noch bedingt Aussagekraft haben.

Die Interessen von Forschung und Politik sind also nicht deckungsgleich. In diesem Austauschverhältnis sind unrealistische und falsche Erwartungen sowie Missverständnisse auf beiden Seiten nicht auszuschließen, können aber verringert werden. Die Wissenschaft kann hier einen wichtigen Beitrag leisten. Vorrangig ist die Sicherung der Qualität wissenschaftlicher Politikberatung, für die wiederum wesentlich ist, dass Beratungsleistungen sowohl den Anforderungen der Wissenschaft als auch den Erfordernissen der Politik gerecht werden.⁴ Eine Beratungsleistung erfüllt die Anforderungen der Forschung, wenn das zu transferierende Wissen durch empirisch belastbare Daten, wissenschaftlich einschlägige Methoden und akademische Diskussionen so weit abgesichert ist, dass verlässliche Aussagen über konkrete politische Fragen getroffen werden können. Hierzu bedarf es einer Forschung, die dem „state of the art“ genügt, sich an internationalen Standards der evidenzbasierten Politikberatung ausrichtet und den „Regeln der guten wissenschaftlichen Praxis“ genügt.⁵

3 MAUSE, HEINE 2003.

4 BERLIN-BRANDENBURGISCHE AKADEMIE DER WISSENSCHAFTEN 2008.

5 DEUTSCHE FORSCHUNGSGEMEINSCHAFT 2013.

Den Erfordernissen der Politik wird das Beratungswissen gerecht, wenn es anschlussfähig an das Wissen und die Wahrnehmung politischer Akteure ist und damit von diesen grundsätzlich akzeptiert wird und umgesetzt werden kann.⁶ Ritter (1982) nennt für eine ergebnisoffene, unabhängige und damit politisch nicht opportune wissenschaftliche Politikberatung verschiedene Qualitätskriterien. Hierzu zählen die Unabhängigkeit des Beraters bzw. der Beraterin, eine gewisse Rotation von Beratern und zu Beratenden, die Vermeidung von Interessenskonflikten im Kontext der Beratungssituation und die Veröffentlichung der Ergebnisse.

Eine gute wissenschaftliche Politikberatung verfolgt die Strategie, die beiden Bezugssysteme soweit wie möglich auszubalancieren.⁷ Dazu gehört freilich auch das Eingeständnis von Politik und Praxis, nicht schon alles zu wissen und bewusst eine Kultur der Evaluation sowie der Korrekturmöglichkeiten von Maßnahmen zuzulassen. Natürlich darf sich die Wissenschaft dabei nicht überhöhen. Sie hat nicht mehr und nicht weniger als eine beratende Funktion. Auch auf Seiten der Forschung muss es die Einsicht geben, dass für gute (politische) Entscheidungen ergänzend auch Erkenntnisse von hoher Bedeutung sein können, die auf andere Art und Weise generiert werden, z.B. auf der Basis von relevantem Praxiswissen.

Eine evidenzbasierte Politikberatung hat prinzipiell eine ganze Reihe von Ansatzpunkten. Typischerweise setzt sie rechtzeitig vor der Entscheidung über die Einführung von Maßnahmen ein und hat nach deren Umsetzung einen langen Atem. Eine konsequente wissenschaftliche Begleitung institutioneller Reformen umfasst somit neben der Ex-ante-Evaluation auch die Ex-post-Evaluation.

Im Vorfeld politischer Interventionen ist wissenschaftliche Politikberatung zunächst vielfach auf Sekundärstudien angewiesen. Das so generierte Beratungswissen ist freilich davon abhängig, wie aktiv die Forschung in dem betreffenden Themenfeld war. Darüber hinaus gibt es aber selbst vor wichtigen institutionellen Änderungen eine ganze Reihe von Möglichkeiten der Primärforschung, die als Ex-ante-Evaluation deklariert werden können. Hier stehen der Wissenschaft –

6 BÄHR 2018.

7 WEINGART, LENTSCH 2008.

je nach Fragestellung – verschiedene Optionen zur Verfügung, die auch nebeneinander zum Einsatz kommen können. In Betracht kommen u.a. Labor- und Feldexperimente sowie modellgestützte Simulationen und Szenarioanalysen.

Nach Umsetzung der Maßnahmen konzentriert sich die wissenschaftliche Begleitung üblicherweise auf Untersuchungen, die auf eine kontrafaktische Situation abstellen. Bei solchen Ex-post-Evaluationen steht die Frage im Vordergrund: Was wäre ohne die politische Intervention passiert? Für deren Durchführung bedarf es zunächst einer Definition der Wirkungsziele sowie einer Identifikation möglicher Nebenwirkungen. Um die kontrafaktische Situation abzubilden, bedarf es einer geeigneten Kontrollgruppe. Die Auswahl der passenden Vergleichsgruppe geschieht in der Praxis vor allem über statistische Matching-Methoden, mit deren Hilfe sogenannte „statistische Zwillinge“ der Teilnehmenden bestimmt werden. Ex-post-Evaluationen können sich auf Makroeffekte (z.B. durch die Ausnutzung regionaler Variation), Mikroeffekte (bezogen auf Betriebe oder Personen) und ergänzend (ohne Vergleichsgruppen) auf die Implementation von Maßnahmen beziehen. Dabei geht es nicht allein um kurzfristige Wirkungsanalysen, sondern auch und gerade um die Untersuchung der Stabilität der Wirkungen im Zeitablauf. Von Interesse sind nicht nur direkte Effekte, sondern auch indirekte Effekte im Sinne von erwünschten und nicht erwünschten Nebenwirkungen.

3. Gründe für die hohe Intensität evidenzbasierter Politikberatung im Politikfeld Arbeitsmarkt

Das Politikfeld Arbeitsmarkt ist von gesellschaftspolitischem Interesse, weil es um die Lebenschancen von Menschen geht und der Grad der Beschäftigung von immenser volkswirtschaftlicher Bedeutung ist. Von jeher ist der Arbeitsmarkt Gegenstand politischer Kontroversen, weil es zwischen den Hauptakteuren, nämlich Arbeitgebern und Beschäftigten, nicht zu leugnende Interessenskonflikte gibt.

Aktuell kann dies am Beispiel des Diskurses zu einer verstärkten Förderung der beruflichen Weiterbildung von Arbeitnehmerinnen und Arbeitnehmern (Stichwort: „Qualifizierungsoffensive“) illustriert werden. Das Bundesministerium für Arbeit und Soziales begründet sein umfassendes Maßnahmenpaket folgendermaßen: „Deswegen ist die zentrale

Antwort auf den vor allem digital getriebenen Strukturwandel die Qualifizierung von Arbeitnehmerinnen und Arbeitnehmern. [...] Gemeinsam müssen wir dafür sorgen, dass alle Arbeitnehmerinnen und Arbeitnehmer die Chance haben, sich die richtigen Kompetenzen und Qualifikationen anzueignen, um sich sicher auf dem Arbeitsmarkt zu bewegen. Auch im Interesse von Fachkräftesicherung gilt es, Qualifikationen durch Fortbildungen zu erneuern sowie berufliche Aufstiege – oder wenn nötig – auch Umstiege zu ermöglichen.“⁸

Die Bundesvereinigung der deutschen Arbeitgeberverbände (BdA) kritisiert die Reichweite der Maßnahmen und führt dazu aus: „Die Vorschläge zu einer sog. Qualifizierungsoffensive aus dem Bundesarbeitsministerium sind unausgegoren, teuer und tragen nicht zur besseren Integration in den Arbeitsmarkt bei. Im Gegenteil: Die Veränderung von Förderleistungen nach dem Gießkannen-Prinzip geht an vielen vorbei, die wirklich Unterstützung benötigen.“⁹

Diese für den Arbeitsmarktkontext beispielhafte Kontroverse kann auf eine sachliche Grundlage gestellt werden, wenn einschlägige Forschungsevidenz einbezogen wird. Dies gilt bereits vor der Einführung durch die Nutzbarmachung des vorhandenen Wissens und geeigneter Instrumente der Ex-ante-Evaluation wie auch nach deren Inkraftsetzung durch eine umfassende Begleitforschung. Ohne die Einbindung einer evidenzbasierten Beratung mit tragfähigen wissenschaftlichen Befunden droht der Politik ein Blindflug, der zu folgenreichen Fehlentscheidungen führen kann.

Diese Lektion lernte die Arbeitsmarktpolitik bereits in den 1990er Jahren und der ersten Dekade des neuen Jahrtausends. Zu dieser Zeit spielten Arbeitsbeschaffungsmaßnahmen (ABM) noch eine große Rolle. Diese zur damaligen Zeit massiv eingesetzte, sozialversicherungspflichtige Form öffentlich geförderter Beschäftigung sollte angesichts von Massenarbeitslosigkeit und insbesondere der Transformation in Ostdeutschland als Marktersatz fungieren und möglichst Brücken in den ersten Arbeitsmarkt schlagen. Im politischen Raum wurden immer wieder Zweifel laut, ob die so definierten Ziele erreicht werden können. Auch unerwünschte Nebeneffekte wie die Verdrängung regulärer Beschäftigung durch ABM wurden kontrovers diskutiert.

8 BUNDESMINISTERIUM FÜR ARBEIT UND SOZIALES 2018.

9 BDA DIE ARBEITGEBER 2018.

Vor diesem Hintergrund wurden Forschungsprojekte zu ABM aufgesetzt, die für die Wirkungsforschung hierzulande als wegweisend betrachtet werden können.¹⁰ Die Studien wiesen für die Fördereintritte sogenannte „Einbindungseffekte“ nach. Während der Förderung suchten ABM-Teilnehmerinnen und -Teilnehmer offenbar weniger intensiv nach Erwerbsarbeit. Hierdurch ergaben sich weit geringere Eingliederungschancen für Teilnehmende als für Vergleichspersonen. Nur bestimmte Gruppen, wie z.B. Langzeitarbeitslose, konnten von diesen Maßnahmen profitieren. ABM zeigten danach im Vergleich zu anderen Instrumenten der Arbeitsmarktpolitik, wie z.B. Eingliederungszuschüssen oder Gründungsförderung, die am wenigsten vorteilhaften Ergebnisse auf der individuellen Ebene.

Die Evaluation der ABM kann in Verbindung mit der Wirkungsforschung zur beruflichen Weiterbildung für Arbeitslose als so etwas wie der Startschuss einer sich in der Arbeitsmarktpolitik entwickelnden Evaluationskultur gesehen werden.¹¹ Als endgültiger Durchbruch muss in diesem Kontext aber die umfassende Evaluation der sog. „Hartz-Reformen“ gesehen werden. Diese zwischen 2003 und 2005 umgesetzte große Arbeitsmarktreform enthielt Maßnahmen der Arbeitsmarktpolitik (wie z.B. Änderungen der Gründungsförderung), des Arbeitsrechts (wie z.B. der Neuregelung der Leiharbeit), der Gewährung von Leistungen (z.B. der Einführung von Hartz IV) und zur Verbesserung der Arbeitsmarktdienstleistungen (u.a. mit einer Reform der Bundesagentur für Arbeit). Mit einem bisher in der Arbeitsmarktforschung nicht gekannten Mittelaufwand beauftragte das Bundesministerium für Arbeit und Soziales (BMAS) damals eine Vielzahl von Forschungsinstituten und Universitäten, die verschiedene Einzelmaßnahmen genauer unter die Lupe nahmen. Hierzu wurde ein multidisziplinärer Ansatz gewählt, der inzwischen als Standard in der Maßnahmenevaluation gilt. Darin werden spezifische Methoden der qualitativen empirischen Wirtschafts- und Sozialforschung (wie z.B. Implementationsanalysen) mit umfangreichen deskriptiven Auswertungen sowie ökonometrischen Analysen mit dem Ziel der Identifikation kausaler Zusammenhänge kombiniert.¹²

Seitdem sind viele Gesetze oder Gesetzesänderungen im Politikfeld Arbeitsmarkt durch die Forschung begleitet worden oder sogar mit Evalu-

10 CALIENDO ET AL. 2004; LECHNER, WUNSCH 2009.

11 Zur Evaluation der beruflichen Weiterbildung: u.a. FITZENBERGER ET AL. 2008.

12 HEYER ET AL. 2012.

ationsklauseln versehen worden. Zudem erhielt das Institut für Arbeitsmarkt- und Berufsforschung (IAB) im Zuge der Hartz-Reformen explizit den gesetzlichen Auftrag, die Arbeitsmarktpolitik sowohl im Bereich der Arbeitslosenversicherung als auch die Leistungen der Grundsicherung regelmäßig zu evaluieren.

4. Bedingungen für eine gute Politikberatung

Die im Bereich Arbeitsmarkt etablierte Wirkungsforschung ist jedoch alles andere als voraussetzungslos. Eine erste wesentliche Voraussetzung für eine evidenzbasierte Politikberatung besteht darin, dass Forscherinnen und Forscher unabhängig sind. Dies schließt selbstverständlich nicht einen intensiven Dialog der Wissenschaft mit den Entscheidungsträgern in Politik und Praxis aus. Im Gegenteil: Evidenzbasierte Politikberatung profitiert von einem solchen Dialog. Hierdurch wächst auf der Seite der Forschung das Verständnis für Fragestellungen von Politik und Praxis. Auf der Seite von Politik und Praxis entwickelt sich durch den Austausch ein Verständnis für die Möglichkeiten und Grenzen evidenzbasierter Politikberatung. Insbesondere geht es um die Erkenntnis, dass eine exzellente Wirkungsforschung Zeit benötigt, um valide Aussagen treffen zu können.

Politik und Praxis tun darüber hinaus gut daran, bei Reformen und Modellversuchen die Forschung frühzeitig einzubinden. Dies ermöglicht nicht nur, bereits vorhandenes Beratungswissen zeitig nutzbar zu machen. Die Einbeziehung erleichtert es auch, gute Bedingungen für eine exzellente Wirkungsforschung herzustellen. So können bei Modellversuchen schon im Vorfeld die Voraussetzungen für randomisierte Feldexperimente mitgedacht werden. Auch bei wissenschaftlicher Begleitforschung zu Reformen kann es vor Einführung den Bedarf einer „Nullmessung“, z.B. durch eine Befragung, geben. Genauso ist die Bildung geeigneter Kontrollgruppen bei der Implementation von Maßnahmen zu berücksichtigen. Bei neuen Themen müssen ggf. zuerst neue Daten generiert werden oder auch Messkonzepte für neue Zielgrößen entwickelt werden.

Von entscheidender Bedeutung für evidenzbasierte Politikberatung sind qualitativ hochwertige Daten. Je nach Politikfeld kommen unterschiedliche Datenquellen (z.B. Prozess- und Erhebungsdaten) für verschiedene Einheiten (z.B. Betriebe, Haushalte und Individuen) in Betracht. Durch den Fokus auf die Eingliederungschancen sind in der

Arbeitsmarktpolitik individuelle Erwerbsverläufe von besonderem Interesse. So kann den Wirkungen von Interventionen nachgegangen werden und es können den Teilnehmerinnen und Teilnehmern vergleichbare Pendants gegenübergestellt werden. Vorteilhaft ist dabei der Datenzugang für die einschlägige Forschung (z.B. durch Forschungsdatenzentren), weil hierdurch Studien replizierbar sind und durch alternative Analysestrategien der Wissensstand langfristig erweitert werden kann.¹³

Um eine evidenzbasierte Politikberatung dauerhaft zu etablieren, ist zudem eine ausreichende Finanzierung der Wirkungsforschung zu gewährleisten. Eine Stetigkeit der Förderung hilft beim Aufbau der Analysekompetenz von Wissenschaftlerinnen und Wissenschaftlern auf dem Feld der Maßnahmenevaluation. Im Politikfeld Arbeitsmarkt steht mit dem IAB seit mehr als 50 Jahren ein zentrales Forschungsinstitut zur Verfügung (siehe Kasten), das kontinuierlich evidenzbasierte Politikberatung betreibt. Darüber hinaus sind die Einrichtung von Graduiertenprogrammen, Nachwuchsgruppen und Stiftungsprofessuren an Hochschulen und Universitäten geeignete Wege, um die Qualität der Wirkungsforschung zu steigern und zu sichern.

Exkurs: IAB als institutioneller Evaluierer

Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) ist eine besondere Dienststelle der Bundesagentur für Arbeit (BA). Die Unabhängigkeit des IAB ist formal abgesichert. Hierzu trägt die räumliche, organisatorische und personelle Trennung vom Verwaltungsbereich der BA bei. Seitens des Gesetzgebers hat das IAB u.a. den Auftrag, die Arbeitsmarktpolitik im Bereich der Arbeitslosenversicherung (§ 282 SGB III) und die Leistungen der Grundsicherung (§ 55 SGB II) zu evaluieren. Kernaufgaben des IAB sind Forschung, (wissenschaftliche) Beratung und Datenproduktion. Das IAB zielt darauf, mit exzellenter Forschung (insbesondere empirische) Evidenz für Politik und Praxis zu liefern. Dies setzt ein hohes Maß an Relevanz der Forschung voraus. Das IAB profitiert von der Nähe zu den arbeitsmarktpoliti-

¹³ KIRCHGÄSSNER 2013.

schen Entscheidungsträgern in der BA mit ihrem Verwaltungsrat und dem Bundesministerium für Arbeit und Soziales (BMAS). Durch den intensiven Dialog können nicht nur Forschungsergebnisse eingebracht werden und in politisches Handeln einfließen, sondern auch Fragestellungen für die Forschung kontinuierlich und frühzeitig generiert werden. Durch den intensiven Austausch mit den Akteuren hat das IAB ein hohes Maß an institutionellem Wissen aufgebaut, dass die evidenzbasierte Politikberatung maßgeblich unterstützt. Die dauerhafte institutionelle Verankerung in der BA gewährleistet die Nachhaltigkeit der Themenbearbeitung. Das IAB hat keinen Alleinvertretungsanspruch in der Arbeitsmarktforschung. Im Gegenteil: Es steht zu den Prinzipien des Wettbewerbs und der Kooperation in der Forschung. Das Forschungsdatenzentrum der BA im IAB stellt der Wissenschaftslandschaft evaluationsgeeignete Daten zur Verfügung und berät interessierte Wissenschaftlerinnen und Wissenschaftler bei Bedarf. Die Forschung unterliegt den Regeln der guten wissenschaftlichen Praxis. Das IAB hat sich aufgrund seiner besonderen Verantwortung „Leitlinien guter wissenschaftlicher Politikberatung“ gegeben. Darin werden vier zentrale Elemente genannt: qualitätsgesicherte wissenschaftliche Befunde, politische Neutralität bei der Beratung, adressatengerechte Aufbereitung der Forschungsbefunde sowie Transparenz über die Beratungsinhalte in Form von Publikationen. Im Einklang damit steht auch das Drittmittelkonzept des IAB. Darin hat sich das IAB (selbst) verpflichtet, nur Projekte anzunehmen, bei denen die Befunde – nach einer gewissen Karenzzeit – unweigerlich veröffentlicht werden.

Evidenzbasierte Politikberatung ist methodisch und wissenschaftlich anspruchsvoll. Sie setzt auf der Seite der Adressaten entsprechende Kenntnisse voraus, um die Erkenntnisse nachvollziehen zu können und für das Tagesgeschäft nutzbar zu machen. Damit kommt auch der Frage der Weiterbildung der Adressaten eine große Bedeutung zu. Unabhängig davon ist die Wissenschaft selbst gefragt, die Ergebnisse der Studien so darzustellen, dass sie auch für die interessierte Fachöffentlichkeit gut verständlich sind.

Schließlich gehört zu einer etablierten Evaluationskultur auch, dass Politik und Praxis mit unbequemen Forschungsergebnissen leben können

und mit diesen in der Öffentlichkeit auch konfrontiert werden wollen. Denn durch evidenzbasierte Politikberatung verfügen Entscheidungsträger zwar über mehr Befunde für die politische Gestaltung, sie stehen aber gleichzeitig unter größerem Rechtfertigungsdruck, insbesondere dann, wenn sich das politische Handeln über die Forschungsergebnisse hinwegsetzt.

5. Fazit

Eine evidenzbasierte Politikberatung stiftet in allen Politikbereichen Nutzen, weil Entscheidungsträgern damit eine sachliche Grundlage für Politikgestaltung zur Verfügung steht und sich damit das Risiko möglicherweise kostspieliger Fehlentscheidungen reduzieren kann. Im Politikfeld Arbeitsmarkt hat sich in den letzten beiden Dekaden eine „Evaluationenkultur“ entwickelt, die durch die Etablierung eines spezialisierten Forschungsinstituts (IAB), eines exzellenten Datenzugangs für die inländische und ausländische Forschung und einer aktiven arbeitsmarktpolitischen Forschung an Universitäten und außeruniversitären Forschungseinrichtungen geprägt wird. Im Bereich der Ex-post-Evaluation sind durch eine Vielzahl von Studien erhebliche Erkenntnisfortschritte zu verzeichnen, die der Politik als wichtige Orientierungshilfe dienen. Es ist jedoch zu erwarten, dass zukünftig angesichts der absehbaren strukturellen Umbrüche und damit einhergehender, teils fundamentaler Umsteuerungsbedarfe der Ex-ante-Evaluation eine noch größere Rolle zukommen kann. Die Arbeitsmarktforschung ist an dieser Stelle bereits gut aufgestellt. Die wissenschaftliche Abschätzung von Politikfolgen vor der Einführung einschneidender Maßnahmen kann helfen, Fehlsteuerungen zu vermeiden.

Literatur

Bähr, H. (2017). „Die Qualität wissenschaftlicher Politikberatung. Unbestimmtheit durch wissenschaftliche Unsicherheit und politische Mehrdeutigkeit“. *Leviathan* 45:3, 404-427.

Bähr, H. (2018). „Wissenschaftliche Politikberatung: Das IAB setzt auf fünf Qualitätsprinzipien“. *IAB-Forum*.

BDA DIE ARBEITGEBER (2018). „Stellungnahme zur Qualifizierungsoffensive Wissen und Sicherheit für den Wandel des Bundesarbeitsministeriums“. Berlin, 30. Mai 2018.

Berlin-Brandenburgische Akademie der Wissenschaften (2008). „Leitlinien Politikberatung“. Berlin.

Buchanan, J.M., Tullock, G. (1962). „The Calculus of Consent. Logical Foundations of Constitutional Democracy.“ Ann Arbor: University of Michigan Press.

Bundesministerium für Arbeit und Soziales (2018). „Qualifizierungsoffensive Wissen und Sicherheit für den Wandel“. Berlin, 30. Mai 2018.

Caliendo, M., Hujer, R., Thomsen, S.L. (2004). „Evaluation der Eingliederungseffekte von Arbeitsbeschaffungsmaßnahmen in reguläre Beschäftigung für Teilnehmer in Deutschland“. Zeitschrift für ArbeitsmarktForschung 37:3, 211-237.

Dehling, J., Schubert, K. (2011). „Ökonomische Theorien der Politik“. Wiesbaden: VS Verlag für Sozialwissenschaften.

Deutsche Forschungsgemeinschaft (2013). „Sicherung guter wissenschaftlicher Praxis. Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission ‚Selbstkontrolle in der Wissenschaft‘“. Weinheim: Wiley-VCH Verlag.

Downs, A. (1957). „An Economic Theory of Democracy“. New York: Harper.

Fitzenberger, B., Osikominu, A., Völter, R. (2008). „Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany“. Annales d'Economie et de Statistique 91-92, 321-355.

Heyer, G., Koch, S., Stephan, G., Wolff, J. (2012). „Evaluation der aktiven Arbeitsmarktpolitik. Ein Sachstandsbericht für die Instrumentenreform 2011“. Zeitschrift für ArbeitsmarktForschung 45:1, 41–62.

Kirchgässner, G. (2013). „Zur Rolle der Ökonometrie in der wissenschaftlichen Politikberatung“. Perspektiven der Wirtschaftspolitik 14:1-2, 3-30.

Lechner, M., Wunsch, C. (2009). „Are Training Programs More Effective When Unemployment is High?“ Journal of Labor Economics 27:4, 653-692.

Mause, K., Heine, K. (2003). „Ökonomische Analysen wissenschaftlicher Politikberatung“. Politische Vierteljahresschrift 44:3, 395–410.

Ritter, E.-H. (1982). „Perspektiven für die wissenschaftliche Politikberatung? Beobachtungen aus der Sicht der Praxis“. Politische Vierteljahresschrift, Sonderheft 13, 458-464.

Weingart, P., Lentsch, J. (2008). „Wissen – Beraten – Entscheiden. Form und Funktion wissenschaftlicher Politikberatung in Deutschland“. Weilerswist: Velbrück Wissenschaft.

Zimmermann, K.F. (2011). „Evidenzbasierte Politikberatung“. Vierteljahreshefte zur Wirtschaftsforschung 80:1, 23-33.

Zimmermann, K.F. (2014). „Evidenzbasierte wissenschaftliche Politikberatung“. IZA Standpunkte 74.

Evidenzbasierte Beratung in der Rentenpolitik

Von AXEL BÖRSCH-SUPAN

Korrespondenzadresse:

Max-Planck-Institut für Sozialrecht und Sozialpolitik (MPISOC)

Munich Center for the Economics of Aging (MEA)

Amalienstr. 33

80799 München

boersch-supan@mea.mpisoc.mpg.de

1. Einleitung

In fast allen OECD-Ländern steht die Reform der Rentensysteme ganz oben auf der sozialpolitischen Agenda. Auslöser der Debatte ist der sich beschleunigende demographische Wandel, der den Druck auf die Beiträge und das Rentenniveau der umlagefinanzierten Rentensysteme massiv erhöht. Während über die Notwendigkeit einer Reform der Alterssicherungssysteme fast überall Einigkeit herrscht, ist deren Ausgestaltung höchst umstritten. Die zentrale Frage dieses Beitrags ist, inwieweit wissenschaftliche Evidenz der Politik in der Vergangenheit geholfen hat und in der Zukunft helfen kann, diese Streitfragen zu klären.

Der Kranz der politischen und akademischen Fragen rund um die Rentenreform ist ein Musterbeispiel, wie in der modernen Volkswirtschaftslehre Theorie und Empirie ineinandergreifen könnten, um der Politik Handlungsempfehlungen nahezulegen. Unsere viel gescholtene Wissenschaft ist dort stark, wo es eine klare Fragestellung, eine robuste Theorie und gute Daten gibt. Dies ist in diesem Gebiet der „Altersökonomik“ klar der Fall. Allerdings können auch hier Effizienz- und Umverteilungsfragen nicht immer getrennt werden. Zumindest ist es aber möglich, die Wirkung auch verteilungspolitischer Fragen recht präzise vorherzusagen.

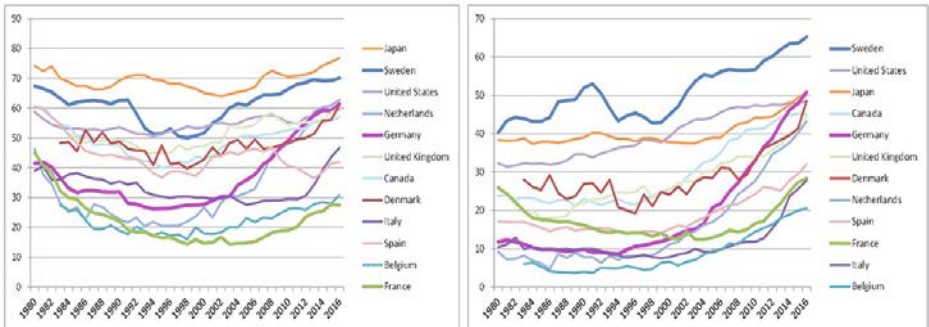
Der Beitrag macht keine tour d’horizon der Rentenpolitik, sondern zeigt am Beispiel des heftig umstrittenen Renteneintrittsalters bzw. der Beschäftigung älterer Arbeitnehmerinnen und Arbeitnehmer, wie im Zusammenspiel von Theorie und Empirie Ursachen und Wirkungen analysiert werden können. Dies hilft der Politik, ex post Maßnahmen zu evaluieren (Abschnitt 2) und ex ante deren Wirkungen abzuschätzen

(Abschnitt 3). Abschnitt 4 zieht ein Fazit, das die diesem Band zugrundeliegenden Leitfragen beantwortet.

2. Ex-post-Evaluation

Abbildung 1 zeigt eine bemerkenswerte Trendumkehr. Bis ca. 2000 ist in den gezeigten OECD-Ländern die Beschäftigung von Männern des Alters 60-64 massiv gesunken. Seit 2000 stieg sie jedoch für Männer und Frauen ebenso massiv wieder an. Vor allem in Deutschland ist dieser Anstieg enorm. Was hat diesen Anstieg bewirkt?

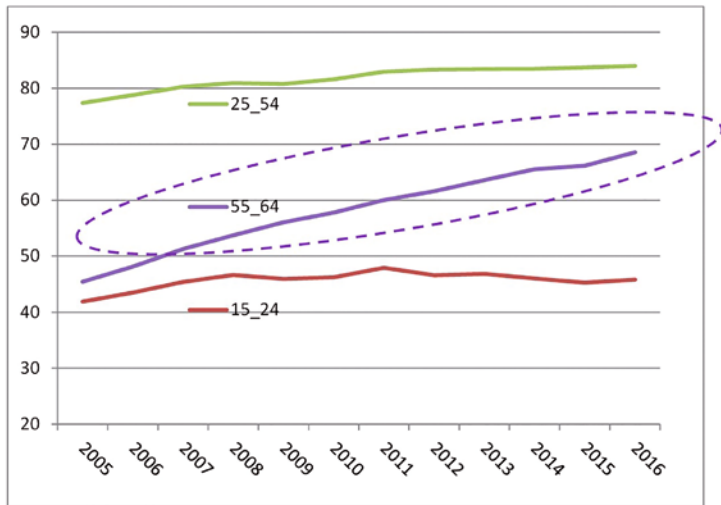
Abbildung 1: Beschäftigung von Männern (links) und Frauen (rechts) des Alters 60-64 in ausgewählten OECD-Ländern



Quelle: OECD. Data extracted on 30 Apr 2018 14:17 UTC (GMT) from OECD.Stat.

Abbildung 2 zeigt, dass die Beschäftigung insgesamt angestiegen ist, was z.B. ein Effekt der Hartz-Reformen gewesen sein mag (vgl. Schmidt in diesem Band). Die Beschäftigung Älterer stieg jedoch deutlich stärker an und kann dadurch nicht erklärt werden.

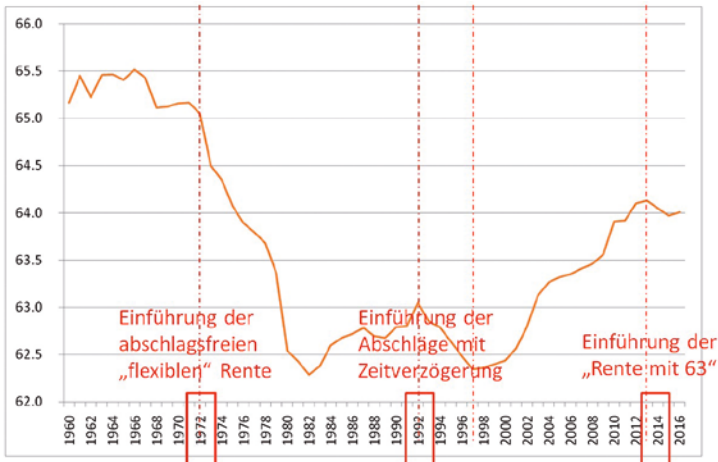
Abbildung 2: Anstieg der Beschäftigung in Deutschland nach Altersklassen



Quelle: OECD. Data extracted on 30 Apr 2018 14:17 UTC (GMT) from OECD.Stat.

Einen Hinweis, dass die Abnahme und der Anstieg der Beschäftigung Älterer keine Wende stetiger Trends war, gibt Abbildung 3, die das durchschnittliche Renteneintrittsalter aufzeigt. Die Abbildung legt nahe, dass das Rentenalter diskontinuierlich insbesondere auf drei Politikeingriffe reagiert hat: die Rentenreform 1972 mit der Einführung der abschlagsfreien „flexiblen Rente“, die Rentenreform 1992, die die rentenrechtlichen Zu- und Abschläge eingeführt hat, und am aktuellen Rand das Rentenpaket 2014 mit der „Rente mit 63“.

Abbildung 3: Durchschnittliches Renteneintrittsalter (Altersrenten für Männer)



Quelle: Deutsche Rentenversicherung

Nun kommt die ökonomische Theorie ins Spiel. Die deutsche gesetzliche Rentenversicherung (GRV) ist in der OECD nicht das krassste, aber ein deutliches Beispiel für ein Rentensystem, das versicherungsmathematisch nicht neutral ist.¹ Die mittleren Barwerte der Rentenauszahlungen, diskontiert und gewichtet mit der Erlebenswahrscheinlichkeit, sind in Deutschland nicht unabhängig vom Renteneintrittsalter. Wer früh in Rente geht, erhält im Barwert eine deutlich höhere Rentenzahlung, während derjenige, der spät in Rente geht, für die kürzeren Laufzeiten nicht ausreichend kompensiert wird. Dies entspricht einer Subventionierung der Frühverrenteten durch die Spätverrenteten und verzerrt das Arbeitsangebot älterer Menschen nach unten.

In der nach 1972 gültigen Regelung gab es bei früherer Pensionierung (Alter 60 bei Frauen, Alter 63 bei Erwerbstätigen mit 35 Jahren Beitragszeit) überhaupt keinen Abschlag. Lediglich die Berechnung der Rente nach Beitragsjahren erhöhte die jährliche Rente um etwa 2,5%, wenn ein Arbeitnehmer oder eine Arbeitnehmerin ein Jahr länger arbeitete. Dies wurde aber dadurch weit mehr als zunichtegemacht, dass diese höhere Rente ein Jahr kürzer bezogen wurde, was bei einer durchschnittlichen Rentenbe-

1 BÖRSCH-SUPAN 2004; QUEISSER, WHITEHOUSE 2006; OECD 2015.

zugsdauer von damals 17 Jahren einer um etwa 6% kürzeren Bezugsdauer entspricht. Der Barwert der Rente sank also mit einer Verschiebung des Renteneintrittsalters. Setzt man diesen Verlust in Relation zu dem Einkommen, das Arbeitnehmerinnen und Arbeitnehmer im Durchschnitt in diesem zusätzlichen Jahr verdienen könnten, berechnet also eine implizite Steuer, die das Rentensystem den Spätverrenteten aufbürdet, erhält man hohe Steuersätze, die bei über 50% nach alter Rentenregelung liegen und immer noch ca. 25% betragen werden, wenn im Jahre 2004 die Regelungen der Rentenreform 1992 vollständig eingeführt worden sind.

Die Rentenreform 1992 führte mit großer Zeitverzögerung Zu- und Abschläge von 3,6% pro Jahr späteren bzw. früheren Renteneintritt ein. Ab dem Jahr 2004 wurde dadurch den Anreizen zur Frühverrentung die Spitze genommen; sie sind bislang jedoch nicht eliminiert, da versicherungsmathematisch neutrale Zu- bzw. Abschläge eher bei 6,5% liegen.² Die ökonomische Theorie ist in dieser Frage eindeutig: Arbeitnehmerinnen und Arbeitnehmer, die diesen Sachverhalt kennen, werden versuchen, zum gesetzlich frühestmöglichen Zeitpunkt in Rente zu gehen, es sei denn, die Arbeit macht ihnen so viel Freude, dass sie dafür zu zahlen bereit sind.

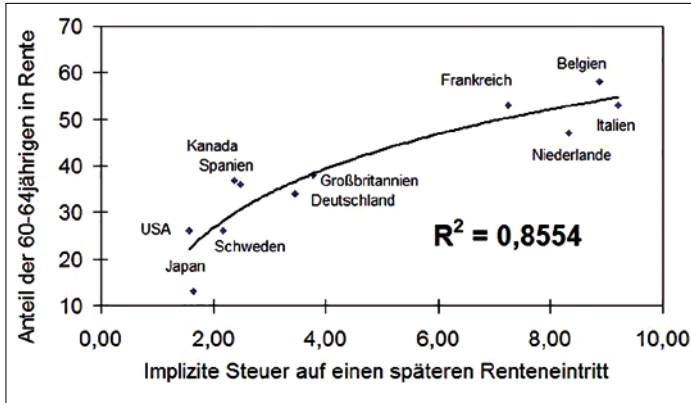
Empirisch erscheint die Lage zunächst weniger klar zu sein, denn zum einen sind die rentenrechtlichen Regeln so komplex, dass dieser Sachverhalt unbekannt sein mag. Zum anderen interessiert Wirtschaftspolitikerinnen und -politiker das konkrete Ausmaß der Reaktion der Erwerbstätigen auf die theoretischen Anzeizeffekte. Tatsächlich ist dies ein Gebiet, auf dem die Empirie sehr viel Information gewinnen konnte. Um das Ergebnis vorwegzunehmen: Die Anreize zur Frühverrentung haben tatsächlich einen starken Einfluss auf das mittlere Rentenalter und damit die Rentenfinanzierung gehabt. So hat z.B. die Einführung der flexiblen Altersgrenze im Jahre 1972 das mittlere Rentenalter um etwa drei Jahre gesenkt, was Mehrkosten von über 15% der Ausgaben verursacht hat.

Es gibt dazu sehr unterschiedliche empirische Belege. Man kann zum ersten aus der internationalen Variation der impliziten Besteuerung der Spätverrenteten lernen. Bei den 11 Ländern, die im Band von Gruber und Wise (1999) betrachtet werden, schwankt dieser implizite Steuersatz stark. Er ist gering in Japan und den USA und besonders hoch in Belgien und Italien. Die Relation zwischen den impliziten Steuersätzen und dem Anteil der

2 WERDING 2007; GASCHÉ 2012.

Erwerbstätigen, die früh in Rente gehen, ist in Abbildung 4 im logarithmischen Maßstab dargestellt. Die hohe Korrelation ist ein deutliches Indiz für die Stärke der Anzeizeffekte, die durch die implizite Besteuerung der Spätereute ausgeübt werden.

Abbildung 4: Effekt der impliziten Besteuerung eines späten Renteneintritts



Quelle: Börsch-Supan 2000.³

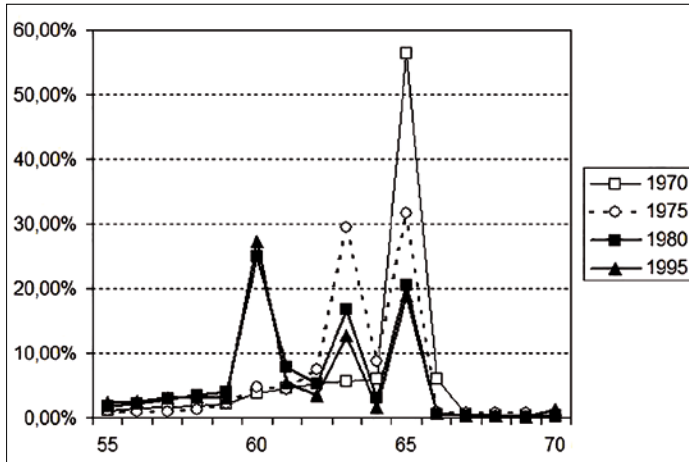
Die kausale Interpretation einer Korrelation aggregierter Daten im Querschnitt ist jedoch angreifbar. Zum einen mag nicht die rentenrechtliche Bevorzugung der Frühverrenteten, sondern die Höhe der Arbeitslosigkeit eine kausale Rolle gespielt haben, die besonders ältere Arbeitnehmerinnen und Arbeitnehmer traf. Allerdings zeigen Riphahn und Schmidt (1997), dass diese Erklärung zeitlich nicht konsistent mit dem Einbruch des Rentenalters war – dieser fand bereits vor dem Anstieg der Massenarbeitslosigkeit statt. Auch das andere oft vorgebrachte Argument, es handele sich nicht um Reaktionen der Arbeitnehmerinnen und Arbeitnehmer, die früh das Rentendasein genießen wollten, sondern um Reaktionen der Arbeitgeber, die das Rentenrecht für Umstrukturierungsmaßnahmen ausnutzen wollten, ist nicht überzeugend – zeigt dies doch gerade, dass das Rentenrecht Anreize zur Frühverrentung bot, die beide Seiten zusammen auf Kosten der übrigen

3 Der implizite Steuersatz wurde über alle Frühverrentungsalter zwischen 60 und 69 Jahren gemittelt. Der Wert 0 entspricht einer Steuer von Null, der Wert 10 einer Steuer von 100% in allen Renteneintrittsaltern.

Beitragszahler ausnutzen konnten.

Ökonometrische Untersuchungen, die dank ihres Quasi-Experimental-characters eher eine kausale Interpretation ermöglichen als Abbildung 4, basieren auf den Optionswertmodellen von Stock und Wise (1980). Sie haben eine hohe Prognosegüte.⁴ In Deutschland sorgte die Rentenreform 1972 für eine Möglichkeit, den Einfluss der geringen Frühverrentungsabschläge wie in einem „natürlichen Experiment“ abzuschätzen. Börsch-Supan (1992) zeigt, dass der damalige Rückgang des Rentenalters keine allmähliche Entwicklung war, sondern eine schnelle Reaktion auf die Einführung der Rentenreform 1972. Insbesondere demonstriert die zeitliche Analyse, dass sich die Rentenzugangsalter bei den frühestmöglichen Altersstufen verlagern, und zwar seit 1972 systematisch vom Alter 65 zum Alter 60, das nun selbst bei männlichen Erwerbstätigen das häufigste Eintrittsalter wurde. Dieser in Abbildung 5 gezeigte zeitliche Zusammenhang lässt eine kausale Interpretation zu.⁵ Diese Arbeiten wurden mehrfach repliziert.⁶

Abbildung 5: Effekt der Rentenreform 1972 auf das Renteneintrittsalter



Quelle: Börsch-Supan, Schnabel 1998.⁷

4 LUMBSDAINE, STOCK, WISE 1992.

5 BÖRSCH-SUPAN, SCHNABEL 1998.

6 SCHMIDT 1995; SIDDIQUI 1997; HANEL 2010.

7 Prozentuale Verteilung des Lebensalters bei Erstbezug einer gesetzlichen Rente aus Versicherung.

Das Modell des Renteneintrittsalters als Funktion der impliziten Besteuerung erklärt auch die in Abbildung 1 gezeigte Trendwende, die der Ausgangspunkt dieses Abschnitts zur Ex-post-Evaluation war. In einer noch unveröffentlichten Studie werten Börsch-Supan und Coile (2018) Paneldaten von 12 Ländern 1980 bis 2015 aus, um den Zusammenhang zwischen der Beschäftigung Älterer und der zunächst ansteigenden und dann wieder zurückgehenden impliziten Besteuerung zu schätzen. Sie erhalten hochgradig signifikante Schätzungen, die das Ab und wieder Auf der Beschäftigung Älterer nachzeichnen.

Als Summe dieser unterschiedlichen empirischen Ansätze lässt sich festhalten, dass sich das mittlere Renteneintrittsalter allein durch eine Erhöhung der Zu- und Abschläge nach oben korrigieren ließe, so dass damit die schwer einsehbare Subventionierung der Frühverrenteten durch die Spätverrenteten beseitigt und gleichzeitig die finanzielle Basis des Umlageverfahrens verbessert werden könnte. Dies ist mit der Reform 1992 teilweise umgesetzt worden. Eine weitere Erhöhung, wie sie in anderen OECD Ländern stattgefunden hat, ist jedoch bislang an politischen Widerständen gescheitert.

3. Ex-ante-Abschätzungen

Das Modell des Renteneintrittsalters als Funktion der impliziten Besteuerung hat sich auch zur Prognose bewährt. Börsch-Supan (2000), Börsch-Supan et al. (2004) und Berkel und Börsch-Supan (2004) haben dieses Modell benutzt, um die Wirkung der Rentenreform 1992 ex ante abzuschätzen. Diese Reform führte, wie bereits erwähnt, graduell Abschläge für die Frührente ein, die für einige Arbeitnehmerinnen und Arbeitnehmer ab 1998, für alle dann ab 2006 griffen. Die Größenordnung der Reaktion auf die Anzeizeffekte wird sehr robust geschätzt und impliziert, dass allein die versicherungsmathematisch zu niedrigen Abschläge der flexiblen Altersgrenze das mittlere Renteneintrittsalter um mindestens zwei Jahre verringert haben. Tatsächlich stieg dann auch das mittlere Rentenalter von 1998 bis 2012 um etwa zwei Jahre an, wie aus Abbildung 3 ersichtlich ist.

Ein zweites Beispiel ist die Einführung der „Rente mit 63“ im Jahr 2014. Die Bundesregierung gab damals die durch diese Maßnahme induzierten zusätzlichen Renteneintritte mit ca. 40.000 pro Jahr an. Börsch-Supan, Coppola und Rausch (2015) schätzten dagegen, dass knapp

200.000 Arbeitnehmerinnen und Arbeitnehmer pro Jahr die neue Regelung in Anspruch nehmen würden. Tatsächlich wurde im Frühjahr 2018 die erste Million erreicht, die mit 63 abschlagsfrei in Rente gehen konnte.

4. Fazit

In Sachen Rentenreform kann ökonomische Theorie und Empirie in vielen Aspekten helfen, Politikmaßnahmen ex post zu evaluieren und deren Wirkung ex ante abzuschätzen. Besonders klar sagt sie uns, dass die Anreize zur Frühverrentung immer noch greifen und zur instabilen Finanzsituation der GRV beitragen.

Diese klaren Aussagen liegen auch an der exzellenten Datenlage und einer großzügigen Förderung empirischer Arbeiten durch die Rentenversicherungen selbst. Die deutsche Rentenversicherung stellt Zeitreihen zum Renteneintritt allen Wissenschaftlerinnen und Wissenschaftlern zur Verfügung. Im Rahmen des Forschungsnetzwerks Alterssicherung (FNA) werden empirische Studien ähnlich finanziell gefördert wie von der US-amerikanischen Social Security Administration, die derzeit z.B. das oben genannte internationale Börsch-Supan-Coile-Projekt finanziert.

Die Wissenschaft wurde bei der Rentenpolitik nicht immer, aber sehr oft einbezogen. Der Sozialbeirat ist vorwiegend mit Wissenschaftlerinnen und Wissenschaftlern besetzt; er liefert das demografische und finanzielle Datengerüst und hat daher einen großen Einfluss auf rentenpolitische Entscheidungen. Auch in die Rentenkommissionen der Bundesregierung 2003, 2014 und 2018 waren immer Wissenschaftlerinnen und Wissenschaftler einbezogen. Die Rentengruppe der „Rürup-Kommission“ 2003 wurde von zwei Wissenschaftlern geleitet. An den Sitzungen des „Rentendialogs“ 2014 nahmen vier Wissenschaftler teil. In der kürzlich gebildeten Rentenkommission „Verlässlicher Generationenvertrag“ sind drei von acht Mitgliedern eine Wissenschaftlerin und zwei Wissenschaftler. Diese institutionelle Einbindung und der damit verbundene enge Austausch zwischen Wissenschaft und Politik erscheint mir sehr wichtig.

Inwieweit die evidenzbasierten Erkenntnisse dann in tatsächliche Politik umgemünzt werden, ist eine andere Frage. Hier gab es wichtige Meilensteine des Erfolgs wie des Scheiterns. Die Rentengruppe der Rürup-Kommission konnte ihre Empfehlungen weitgehend umsetzen (Nachhaltigkeitsfaktor, „Rente mit 67“). Auch einige Anstöße im Rahmen des „Rentendialogs“ (Stärkung der Erwerbsminderungsrente, Be-

triebsrentenstärkungsgesetz) wurden verwirklicht. Die Arbeit der jüngst eingesetzten neuerlichen Rentenkommission hat noch nicht begonnen. Ihre Arbeit ist jedoch nicht leichter geworden dadurch, dass ein gegen wissenschaftlichen Rat konzipiertes Gesetz zur Einführung einer doppelten Haltelinie im Koalitionsvertrag verankert wurde. Auch die „Rente mit 63“ und die „Mütterrente“ wurden gegen wissenschaftlichen Rat verabschiedet. Das Glas ist ebenso halb voll wie es halb leer ist. Populistische Wahlkampfversprechen sind das Haupthindernis, wissenschaftliche Erkenntnisse in die Ausgestaltung der Rentenpolitik einfließen zu lassen.

Literatur

Berkel, B., Börsch-Supan, A. (2004). „Pension Reform in Germany: The Impact on Retirement Decisions“. Finanzarchiv 60, 393-421.

Börsch-Supan, A. (1992). „Population Aging, Social Security Design, and Early Retirement“. Journal of Institutional and Theoretical Economics 148, 533-557.

Börsch-Supan, A., Schmidt, P. (1996). „Early Retirement in East and West Germany“. In: Riphahn, R.T., Snower, D., Zimmermann, K. (Hrsg.), Employment Policy in Transition: The Lessons of German Integration for the Labor Market. Berlin, Heidelberg, Tokyo: Springer.

Börsch-Supan, A., Schnabel, R. (1998). „Social Security and Declining Labor Force Participation in Germany“. American Economic Review 88:2, 173-178.

Börsch-Supan, A. (2000). „Incentive Effects of Social Security on Labor Force Participation: Evidence in Germany and Across Europe“. Journal of Public Economics 78, 25-49.

Börsch-Supan, A. (2004). „Faire Abschlüsse in der gesetzlichen Rentenversicherung“. Sozialer Fortschritt 53, 258-261.

Börsch-Supan, A., Schnabel, R., Kohnz, S., Mastrobuoni, G. (2004). „Micro-Modeling of Retirement Decisions in Germany“. In: Gruber, J., Wise, D. (Hrsg.), Social Security Programs and Retirement Around the World: Micro-Estimation. Chicago: The University of Chicago Press.

Börsch-Supan, A., Coppola, M., Rausch, J. (2015). „Die Rente mit 63: Wer sind die Begünstigten? Was sind die Auswirkungen auf die Gesetzliche Rentenversicherung?“ Perspektiven der Wirtschaftspolitik 16, 264-288.

Gruber, J., Wise, D. (Hrsg.) (1999). „International Comparison of Social Security Systems“. Chicago: The University of Chicago Press.

Hanel, B. (2010). „Financial Incentives to Postpone Retirement and Further Effects on Employment – Evidence from Natural Experiment“. *Labour Economics* 17:3, 1-13.

Lumsdaine, R.L., Stock, J.H., Wise, D.A. (1992). „Three Models of Retirement: Computational Complexity versus Predictive Validity“. In: Wise, D.A. (Hrsg.), *Topics in the Economics of Aging*. Chicago: University of Chicago Press, 16-60.

OECD (2015). „Pensions at a Glance“. Paris: OECD.

OECD (2018). „Employment database“. Paris: OECD. (<http://www.oecd.org/employment/emp/onlineoecdemploymentdatabase.html> [Zugriff März 2018]).

Riphahn, R.T., Schmidt, P. (1997). „Determinanten des Ruhestandes: Lockt der Ruhestand oder drängt der Arbeitsmarkt?“. *Jahrbücher für Wirtschaftswissenschaften* 48:1, 113-147.

Schmidt, P. (1995). „Die Wahl des Rentenalters – Theoretische und empirische Analyse des Rentenzugangsverhaltens in West- und Ostdeutschland“. Frankfurt am Main: Lang.

Siddiqui, S. (1997). „The Pension Incentive to Retire: Empirical Evidence for West Germany“. *Journal of Population Economics* 10:4, 463-86.

Stock, J.H., Wise, D.A. (1990). „The Pension Inducement to Retire: An Option Value Analysis“. In: Wise, D.A. (Hrsg.), *Issues in the Economics of Aging*, 205-230. Chicago: University of Chicago Press.

Queisser, M., Whitehouse, E. (2006). „Neutral or Fair? Actuarial Concepts and Pension-System Design“. *OECD Social, Employment and Migration Working Papers* 40. Paris: OECD.

Werding, M. (2007). „Versicherungsmathematisch korrekte Rentenabschläge für die gesetzliche Rentenversicherung“. *ifo Schnelldienst* 16, 19-32.

Evidenzbasierte Wirtschaftspolitik: Lehren aus dem Politikfeld Arbeitsmarkt

Von BERND FITZENBERGER

Korrespondenzadresse:

Humboldt-Universität zu Berlin

Wirtschaftswissenschaftliche Fakultät

Spandauer Str. 1

10178 Berlin

fitzenbb@hu-berlin.de

Die Evaluation von Maßnahmen der Arbeitsmarktpolitik – und damit die empirische Arbeitsmarktforschung insgesamt – nimmt in den letzten zwei Jahrzehnten eine Vorreiterrolle in der evidenzbasierten Wirtschaftspolitik ein, sowohl international als auch in Deutschland.¹ Arbeitsmarktpolitische Maßnahmen oder Reformen im Arbeitsmarkt werden inzwischen häufig, wenn auch nicht immer, einer wissenschaftlichen Begleitforschung unterzogen, oftmals in Verbindung mit einer empirischen Wirkungsanalyse. Herausgehobene Beispiele für die zunehmende Evidenzbasierung der Arbeitsmarktpolitik in Deutschland sind die Konzeption und Durchführung der Evaluation der Hartz-Reformen sowie die Evaluation der Wirkungen der Branchenmindestlöhne. Eine Institutionalisierung der Evidenzbasierung erfolgt zum Beispiel durch die systematische Bereitstellung von qualitativ hochwertigen, administrativen Arbeitsmarktdaten von Seiten der Bundesagentur für Arbeit und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) sowie des Statistischen Bundesamtes und der Statistischen Landesämter. Weitere Beispiele sind in den 2000er Jahren die Einrichtung zusätzlicher Forschungsabteilungen im IAB, deren Aufgabe die wissenschaftliche Begleitforschung zum Sozialgesetzbuch (SGB) II darstellt, oder im Jahr 2016 die Einrichtung der Mindestlohnkommission, deren Geschäftsstelle mit der wissenschaftlichen Bewertung der Auswirkungen des Mindestlohns beauftragt ist. Die für Evaluationsforschung zum Arbeitsmarkt bereitgestellten administrativen Daten ergänzen die für wissenschaftliche Eva-

1 HECKMAN ET AL. 1999; FITZENBERGER, HUJER 2002.

lationsstudien verfügbaren Umfragedaten (beispielsweise das Sozio-oekonomische Panel, SOEP).

Trotz zahlreicher positiver Entwicklungen, die im Rahmen dieser Ausführungen nicht alle gewürdigt werden können, gibt es gleichwohl weiteren Verbesserungsbedarf im Hinblick auf die Entwicklung einer umfassenden Evaluationskultur im Politikfeld Arbeitsmarkt in Deutschland. Mit den folgenden Punkten gehe ich zunächst im Rückblick auf einige wichtige Schritte hin zu einer stärker evidenzbasierten Arbeitsmarktpolitik und dann auf weiterhin bestehende Herausforderungen ein. Eine umfassende Darstellung aller Aspekte übersteigt den Rahmen dieser Ausführungen. Hierzu sei auf die Literaturverweise in diesem Dokument und die weiteren Stellungnahmen im Rahmen der von der Deutschen Bundesbank und der Nationalen Akademie der Wissenschaften Leopoldina gemeinsam organisierten Veranstaltung „Evaluation von Finanzmarktreformen: Lehren aus anderen Politikbereichen und Disziplinen“ verwiesen.

- 1) Ein wichtiger Katalysator für eine stärker evidenzbasierte Arbeitsmarktpolitik in Deutschland war die hohe Arbeitslosigkeit um die Jahrtausendwende, deren Persistenz zunehmend Zweifel an der Wirksamkeit der aktiven Arbeitsmarktpolitik schürte. Das Interesse einer Evaluation der Wirkung der mit hohen Ausgaben verbundenen Maßnahmen wuchs. Gleichzeitig wurden arbeitsmarktpolitische Reformen gefordert. Die Diskussion wurde zusätzlich genährt durch den sogenannten Vermittlungsskandal der Bundesanstalt für Arbeit zu Beginn der 2000er Jahre. Alle Aktivitäten der Bundesanstalt kamen in der Folge auf den Prüfstand, die Einrichtung wurde umfassend reformiert sowie in Bundesagentur für Arbeit (BA) umbenannt. Die Organisationsreform der BA ging einher mit umfangreichen Reformen der Arbeitsmarktpolitik (Stichwort Hartz-Reformen), für die die damalige Bundesregierung erstmalig eine umfassende Evaluation der Reformwirkungen vorsah. Diese Evaluation erfolgte in zwei Stufen. In der ersten Stufe wurden von zwei Konsortien² Vorstudien zur Entwicklung von Konzepten für eine Gesamtevaluation der Hartz-Reformen vorgelegt. In der zweiten

2 HAGEN, SPERMANN 2004; FERTIG ET AL. 2004.

Stufe wurden einzelne Evaluationsaufträge für unterschiedliche Teile der Reformen vergeben. Die Aufträge orientierten sich an den Vorstudien und umfassten sowohl quantitative, ökonometrische Evaluationen als auch umfangreiche, eher qualitativ orientierte Implementationsstudien. Auftragnehmer waren meist wirtschafts- und sozialwissenschaftliche Forschungsinstitute, oftmals in enger Kooperation mit der Forschung an Hochschulen.

- 2) Die angemessene Wirkungsanalyse von arbeitsmarktpolitischen Maßnahmen nimmt in der arbeitsökonomischen Forschung spätestens seit den 1990er Jahren einen hohen wissenschaftlichen Stellenwert ein.³ Dies lag sowohl daran, dass die statistisch-ökonometrische Methodenentwicklung große Fortschritte gemacht hatte, als auch daran, dass sich gleichzeitig die Verfügbarkeit von hochwertigen Arbeitsmarktdaten (Individualdaten, Paneldaten mit Regionalinformationen, um nur einige Punkte zu benennen) verbesserte, die eine Durchführung von Evaluationsstudien nach dem aktuellen Stand der Forschung ermöglichte.⁴ Das grundsätzliche Evaluationsproblem besteht darin, dass kontrafaktische Werte für die Zielvariablen (bspw. Beschäftigung, Entlohnung) für alternative, nicht beobachtete Szenarien geschätzt werden müssen. Im Fall der Wirkungsanalyse von arbeitsmarktpolitischen Maßnahmen muss für Teilnehmende geschätzt werden, was kontrafaktisch passiert wäre, wenn sie nicht an der Maßnahme teilgenommen hätten. Umgekehrt muss für Nichtteilnehmende geschätzt werden, was kontrafaktisch passiert wäre, wenn sie an der Maßnahme teilgenommen hätten. Um das kontrafaktische Nichtteilnahmeergebnis zu schätzen, müssen Daten für mit den Teilnehmenden vergleichbare Nichtteilnehmende herangezogen werden.
- 3) Die Methoden der statistisch-ökonometrischen Evaluationsforschung stellen hohe Anforderungen an die verfügbaren Daten. Im genannten Beispiel müssen Teilnehmende und Nichtteilnehmende identifiziert werden, die sich in allen relevanten Variablen hinrei-

3 HECKMAN ET AL. 1999.

4 FITZENBERGER, HUIJER 2002.

chend ähnlich sind. Ende der 1990er Jahre wuchs in Deutschland das Bewusstsein, dass Umfragedaten wie das Sozio-oekonomische Panel - trotz ihrer hohen Bedeutung für die empirische Arbeitsmarktforschung - für die Evaluation von arbeitsmarktpolitischen Maßnahmen nicht immer ausreichend sind. Gleichzeitig forderte die Wissenschaft immer stärker die Verfügbarmachung administrativer Individualdaten des Arbeitsmarktes ein. Solche Daten können inzwischen über die Forschungsdatenzentren – insbesondere das Forschungsdatenzentrum der Bundesagentur für Arbeit und des IAB – bereitgestellt werden. Die Politik unternahm in den frühen 2000er Jahren große Anstrengungen, um administrative Arbeitsmarktdaten für die wissenschaftliche Analyse zur Verfügung zu stellen – und hiervon wurde seitdem auch intensiv bei von der Politik in Auftrag gegebenen Evaluationsstudien Gebrauch gemacht.⁵ Juristisch wird der Zugang über das SGB X geregelt. Forschende müssen in einem aufwendigen Antragsverfahren gegenüber dem Bundesministerium für Arbeit und Soziales (BMAS) überzeugend darlegen, dass ihre Analyse auf Basis dieser Daten arbeitsmarktpolitisch oder sozialpolitisch relevant ist und nicht mit Umfragedaten machbar ist.

- 4) Die Fortschritte hin zu einer evidenzbasierten Arbeitsmarktpolitik sind das Ergebnis einer Win-Win-Situation für Politik und Wissenschaft. Das durch die Methodenentwicklung beförderte, gesteigerte wissenschaftliche Interesse an Evaluationsstudien, die Verbesserung der Datenverfügbarkeit und das zunehmende politische Interesse an belastbaren Evaluationsergebnissen führten in den 2000er Jahren dazu, dass das Politikfeld Arbeitsmarkt eine Vorreiterrolle in evidenzbasierter Politikgestaltung einnehmen konnte. Die Ergebnisse der inzwischen zahlreichen Evaluationsstudien zeigen ein sehr differenziertes Bild der Wirksamkeit unterschiedlicher arbeitsmarktpolitischer Maßnahmen. Beispielsweise zeigten Studien auf Basis von detaillierten administrativen Daten in den 2000er

5 Siehe die Vorschläge zur Evaluation der Hartz-Reformen, FERTIG ET AL. 2004 und HAGEN, SPERMANN 2004, oder die Evaluationen der Branchenmindestlöhne, DOERR, FITZENBERGER 2016.

Jahren, dass berufliche Weiterbildung langfristig positive Beschäftigungswirkungen entfaltet.⁶ Gleichwohl bestehen jedoch auch im Politikfeld Arbeitsmarkt weiterhin wichtige Herausforderungen. Auf einige davon gehe ich im Folgenden ein.

- 5) Wissenschaftliche Begleitforschung auf Basis guter Daten und guter Methoden kann belastbare Evidenz im Hinblick auf die Wirkungen von politischen Reformen oder Maßnahmen auf relevante, messbare Zielgrößen geben. Falls die Wirkungen und Kosten in Geldeinheiten gemessen werden können und Konsens zwischen Wissenschaft und Politik darüber besteht, dass für eine Bewertung eine Nutzen-Kosten-Analyse in Geldeinheiten ausreicht (oft liegen jedoch keine individualisierten Kosteninformationen vor), dann kann man sich im Rahmen einer Evaluationsstudie dazu äußern, ob die betreffende politische Reform oder Maßnahme sinnvoll ist. Im Regelfall kann die Evaluationsstudie jedoch nur die Wirkung auf bestimmte Zielgrößen aufzeigen und es bleibt der Politik überlassen, diese Wirkungen im Lichte der Kosten zu bewerten. Eine allumfassende wissenschaftliche Antwort auf die Frage, ob eine politische Reform oder Maßnahme unter wirtschaftlichen und sozialpolitischen Gesichtspunkten sinnvoll ist, ist im Regelfall nicht möglich, da in die politische Entscheidung ebenfalls Werturteile einfließen, die nicht durch die Wissenschaft gefällt werden können. Gleichwohl sind wissenschaftliche Evaluationsergebnisse sehr wichtig, weil sie Evidenz im Hinblick auf die von der Politik unterstellten Wirkungen liefern und damit den politischen Diskurs versachlichen. Eine „Erfolgsgeschichte“ der evidenzbasierten Politikberatung im Politikfeld Arbeitsmarkt ist das starke Zurückfahren der Arbeitsbeschaffungsmaßnahmen, da für diese Maßnahmen keine bedeutsamen positiven Beschäftigungswirkungen nachgewiesen werden konnten.⁷

Eine große Herausforderung stellen von der Politik genannte Zielgrößen dar, die bisher nicht oder nur schwierig zu messen sind. Dieses Problem stellte sich im Rahmen der Hartz-Evaluationen im

6 Zum Beispiel FITZENBERGER ET AL. 2008.

7 HUJER, THOMSEN 2010.

Hinblick auf das von der Politik genannte Ziel der Erhöhung der Beschäftigungsfähigkeit von Arbeitslosen. Zunehmend sehen deshalb Ausschreibungen von Ministerien für Evaluationsprojekte vor, dass im Rahmen der Evaluationsstudie Messkonzepte für die politischen genannten Zielgrößen entwickelt und empirisch umgesetzt werden. In den Evaluationen der Branchenmindestlöhne sollten beispielsweise die Auswirkungen der Mindestlöhne auf den Arbeitnehmerschutz und den Wettbewerb untersucht werden, wobei für diese Ziele im Rahmen der Evaluationsstudien Konzepte zur Operationalisierung entwickelt werden mussten.⁸ Bei der aktuellen politischen Diskussion um einen „sozialen Arbeitsmarkt“, der möglicherweise ähnlich wie Arbeitsbeschaffungsmaßnahmen für Langzeitarbeitslose konzipiert wird, steht nicht die Wirkung auf die zukünftigen Beschäftigungschancen,⁹ sondern auf andere Zielgrößen - wie bspw. soziale Teilhabe - im Mittelpunkt, die es gegebenenfalls zu operationalisieren gilt. Schon die Operationalisierung kann einen wichtigen Beitrag zur Versachlichung der Diskussion liefern.

- 6) Eine fundierte und wirkungsvolle evidenzbasierte Beratung der Wirtschaftspolitik erfordert eine Qualitätssicherung wissenschaftlicher Evaluationsergebnisse. Zum einen muss, wer Evaluationsergebnisse nachfragt, diese kritisch einschätzen können – insbesondere dann, wenn unterschiedliche Analyseverfahren zu unterschiedlichen Ergebnissen führen. Dies erfordert die Kompetenz in der Wirtschaftspolitik (in Ministerien, Parlamenten, Behörden und betroffenen gesellschaftlichen Gruppen), wissenschaftliche Evaluationsergebnisse angemessen einschätzen zu können, insbesondere im Hinblick auf die Qualität der dahinterstehenden wissenschaftlichen Analyse. Das notwendige wissenschaftliche Knowhow muss bei den Auftraggeberinnen und -gebern von Evaluationsstudien bestehen. Dies wird im Politikfeld Arbeitsmarkt durch einen intensiven Austausch zwischen Forschenden und Nachfragern von Evaluationsergebnissen und durch die Einstellung

8 DOERR, FITZENBERGER 2016.

9 HUJER, THOMSEN 2010.

von im Bereich der empirischen Arbeitsmarktforschung sehr gut ausgebildeten Mitarbeiterinnen und Mitarbeitern in Ministerien und Behörden hergestellt.

Wissenschaftlicher Fortschritt lebt vom wissenschaftlichen Diskurs. Gleiches gilt für die Bewertung von Evaluationsstudien. Hier besteht meiner Ansicht nach ein substanzielles Defizit in Deutschland, auch im Politikfeld Arbeitsmarkt. Von der Politik in Auftrag gegebene Evaluationsstudien werden viel zu selten nach den Kriterien des wissenschaftlichen Diskurses kritisch bewertet. In Doerr und Fitzenberger (2016) nehmen wir eine kritische methodische Bewertung der Evaluationsstudien zu den Branchenmindestlöhnen vor. Von Seiten des BMAS als Auftraggeber der Evaluationsstudien wurde eine solche kritische und transparente Diskussion der Evaluationsstudien nicht durchgeführt. Als positives Beispiel möchte ich die damalige Vergabe zweier Studien zur Konzeption der Hartz-Evaluationen nennen.¹⁰ Diese Vorstudien wurden vom BMAS in Auftrag gegeben und intensiv mit der Wissenschaft in transparenter Weise diskutiert, bevor die eigentlichen Evaluationsstudien in Auftrag gegeben wurden. Aus meiner Sicht wäre es sinnvoll, wenn in ähnlicher Weise Ergebnisse von Evaluationsstudien von den Auftraggebern mit weiteren Wissenschaftlerinnen und Wissenschaftlern transparent und kritisch diskutiert würden, um die Belastbarkeit der erzielten Ergebnisse besser einschätzen zu können und zu einer Versachlichung der Debatte beizutragen. Die Verfahren der Qualitätssicherung in der Wissenschaft (transparente Begutachtung von wissenschaftlichen Arbeiten, offene Diskussion auf Konferenzen, wissenschaftliche Beiräte) können hier Vorbild für die Qualitätssicherung von Evaluationsstudien sein. Zur notwendigen Transparenz gehört auch, dass wissenschaftliche Evaluationsstudien nach einer angemessenen, aber möglichst kurzen Karenzfrist öffentlich gemacht werden und wissenschaftliche Replikationsstudien von Evaluationsstudien – auch im Hinblick auf den Zugang zu datenschutzrechtlich sensiblen Daten innerhalb von Forschungsdatenzentren – ermöglicht werden. Dies sollte juristisch abgesichert werden.

10 HAGEN, SPERMANN 2004; FERTIG ET AL. 2004.

- 7) Zu einer umfassenden Evaluationskultur gehört ein systematisches, zeitnahes Monitoring der politisch vorgegebenen Zielgrößen und wichtiger, damit zusammenhängender Indikatoren. Im Politikfeld Arbeitsmarkt ist dies eine Kernaufgabe des IAB. Wenn eine politische Reform oder Maßnahme ins Auge gefasst wird, sollte die Evaluation mitgedacht werden. Denkbar sind die Durchführung von Feldexperimenten vor der flächendeckenden Umsetzung, eine zeitlich verzögerte Einführung in verschiedenen Regionen oder die Verwendung von Übergangsregeln, um kurzfristige kausale Wirkungen einschätzen zu können. Vorstellbar ist auch eine Art Verfallsdatum für politische Reformen oder Maßnahmen, falls bis zu dem Zeitpunkt keine hinreichend positiven Wirkungen auf eine vorab von der Politik festgelegte Zielgröße erkennbar sind. Experimente gelten gemeinhin als der „Goldstandard“ der Evaluationsforschung und es ist sehr begrüßenswert, dass die Bundesagentur für Arbeit in Einzelfällen Feldexperimente für innovative arbeitsmarktpolitische Maßnahmen durchführt. Feldexperimente spielten eine entscheidende Rolle für die Methodenentwicklung in der Evaluationsforschung, da die experimentellen Ergebnisse als Benchmark für statistisch-ökonomische Verfahren für nichtexperimentelle Daten dienen können.¹¹ Jedoch geraten Experimente, die zeitlich verzögerte Einführung in verschiedenen Regionen oder die Verwendung von Übergangsregeln als Mittel des Erkenntnisgewinns für eine evidenzbasierte Wirtschaftspolitik an ihre Grenzen, wenn es um längerfristige Wirkungen von Reformen oder Maßnahmen geht. Ein wichtiges Beispiel hierfür ist die Evaluation der Wirkungen von langen Weiterbildungsprogrammen für Arbeitslose, die kurzfristig mit Beschäftigungsverlusten (Lock-in-Effekten) einhergehen und die erst mittel- bis längerfristig positive Auswirkungen auf Beschäftigung und Verdienste der Teilnehmerinnen und Teilnehmer zeigen können.¹² Eine experimentelle Evaluation der Teilnahme an Weiterbildung stößt auch deshalb an ihre Grenzen, weil experimentell ein Weiterbildungsangebot gemacht werden kann, es aber der Entscheidung der Arbeitslosen obliegt, ob sie ein

11 HECKMAN ET AL. 1999.

12 CARD ET AL. 2017.

experimentell zugewiesenes Angebot annehmen oder ob sie sich bei Nichtangebot eine alternative Weiterbildung auf dem privaten Weiterbildungsmarkt suchen.

Literatur

Card, D., Kluve, J., Weber, A. (2017). „What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations“. *Journal of the European Economic Association* 16:3, 894-931.

Doerr, A., Fitzenberger, B. (2016). „Konzeptionelle Lehren aus der ersten Evaluationsrunde der Branchenmindestlöhne in Deutschland“. *Journal for Labour Market Research* 49:4, 329-347.

Fertig, M., Kluve, J., Schmidt, C.M., Apel, H., Friedrich, W., Hägele, H. (2004). „Die Hartz-Gesetze zur Arbeitsmarktpolitik: Ein umfassendes Evaluationskonzept“. Berlin: Duncker & Humblot.

Fitzenberger B., Hujer R. (2002). „Stand und Perspektiven der Evaluation der aktiven Arbeitsmarktpolitik in Deutschland“. *Perspektiven der Wirtschaftspolitik* 3:2, 139-158.

Fitzenberger, B., Osikominu, A., Völter, R. (2008). „Get Training or Wait? Long Run Employment Effects of Training Programs for the Unemployed in West Germany“. *Annales d'Economie et de Statistique* 91-92, 321-355.

Hagen, T., Spermann, A. (2004). „Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung“. *ZEW Wirtschaftsanalysen. Schriftenreihe des ZEW* 74.

Heckman, J.J., LaLonde, R.J., Smith, J.A. (1999). „The Economics and Econometrics of Active Labor Market Programs“. In: Ashenfelter, O., Card, D. (Hrsg.), *Handbook of Labor Economics* 3, 1865-2097.

Hujer, R., Thomsen, S.L. (2010). „How Do the Employment Effects of Job Creation Schemes Differ With Respect to the Foregoing Unemployment Duration?“. *Labour Economics*, 17:1, 38-51.

3. Politikfeld Finanzmärkte und Fiskalpolitik

Evidenzbasierte Finanzmarktregulierung – Implikationen aus Sicht der Finanzaufsicht

Von THORSTEN FREIHUBE und FELIX HUFELD

Korrespondenzadresse:

Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin)

Graurheindorfer Str. 108

53117 Bonn

thorsten.freihube@bafin.de

Evidenzbasierte Politikgestaltung ist ein zentrales Element in allen politischen Bereichen. Gerade bei der Regulierung von Finanzmärkten und Finanzintermediären ist der Austausch mit Marktteilnehmern und der Wissenschaft enorm wichtig. Über die von ihnen erhältliche Evidenz, sei sie empirisch, experimentell oder qualitativ, lassen sich wesentliche Erkenntnisse über Anreize und Verhaltensweisen gewinnen. Vor allem die Disziplin Wirtschaftswissenschaft bietet als Sozialwissenschaft umfangreiche Möglichkeiten, Daten von den Märkten und beaufsichtigten Marktteilnehmern, aber auch die Auswertung bereits ergriffener Regulierungsmaßnahmen zu verwenden.

Das Verhalten von Wirtschaftssubjekten wird laut wirtschaftswissenschaftlicher Theorie meist durch deren persönliche Nutzenfunktion bestimmt.¹ Für den Finanzsektor liegt es nahe, die Nutzenfunktion auf eine monetäre Auszahlungsfunktion zu reduzieren, obwohl auch die Akteure auf den Finanzmärkten nicht frei von anderen persönlichen Nutzen wie Macht und Eitelkeit agieren.

Die Europäische Kommission und die deutsche Bundesregierung haben sich den Grundsätzen der „Besseren Rechtsetzung“ (*Better*

1 STEPHAN, ALHEIM 1996.

Regulation)² verschrieben. Um bessere Ergebnisse zu erzielen, soll sich die Rechtsetzung auf Fakten und einen transparenten Prozess stützen. Dabei sollen Bürgerinnen und Bürger und Interessenträger (z.B. Unternehmen, öffentliche Verwaltungen und Wissenschaftskreise) in den gesamten Verlauf einbezogen werden. Im Vorfeld der Rechtsetzung sollen die Gesetzesfolgen abgeschätzt werden. Die Gesetzesfolgenabschätzung (Regulatory Impact Assessment) stellt das Kerninstrument der Better-Regulation-Initiative dar. In der Gemeinsamen Geschäftsordnung der Bundesministerien (GGO) findet sich eine Definition von Gesetzesfolgen in §44 (1):

„Unter Gesetzesfolgen sind die wesentlichen Auswirkungen des Gesetzes zu verstehen. Sie umfassen die beabsichtigten Wirkungen und die unbeabsichtigten Nebenwirkungen. Die Darstellung der voraussichtlichen Gesetzesfolgen muss im Benehmen mit den jeweils fachlich zuständigen Bundesministerien erfolgen und hinsichtlich der finanziellen Auswirkungen erkennen lassen, worauf die Berechnungen oder die Annahmen beruhen. Es ist darzustellen, ob die Wirkungen des Vorhabens einer nachhaltigen Entwicklung entsprechen, insbesondere welche langfristigen Wirkungen das Vorhaben hat. (...)“³

Die Level-3-Komitees im Lamfalussy-Verfahren, Committee of European Banking Supervisors (CEBS), Committee of European Securities Regulators (CESR) und Committee of European Insurance and Occupational Pensions Supervisors (CEIOPS), haben bereits 2008 in ihrem Rahmenwerk „Regulatory Impact Assessment Guidelines“ (RIA)⁴ festgelegt, wie die übergeordneten Ziele der Regulierung unter Berücksichtigung vorhandener Evidenz mit verschiedenen Regulierungsoptionen erreicht werden können. Eine entsprechende Regelung findet sich auch in der Rechtsgrundlage der Europäischen Aufsichtsbehörden (European Su-

2 EUROPÄISCHE KOMMISSION o.J.

3 Vgl. § 44 Gemeinsame Geschäftsordnung der Bundesministerien.

4 IMPACT ASSESSMENT GUIDELINES FOR EU LAMFALUSSY LEVEL 3 COMMITTEES 2008. Ref: CESR/07-089b , Ref: CEBS 2007 58 , Ref: CEIOPS-3L3-08/06

pervisory Authorities – ESAs), welche aus den 3L3-Komitees hervorgegangen sind. Die ökonomische Analyse im Rahmen eines Regulatory Impact Assessments (RIA) beschränkt sich aber keineswegs auf die Level-3-Arbeiten (im Europäischen Lamfalussy Regulierungsverfahren),⁵ sondern ist auf allen Ebenen des Regulierungsprozesses erforderlich.⁶

Am Anfang (insbesondere auf Level 1, also der Ebene von Richtlinien und Verordnungen) eines Regulatory Impact Assessments (RIAs) steht die Identifizierung des Problems. Bereits an dieser Stelle hilft Evidenz, das Problem zu beschreiben und insbesondere aufzuzeigen, ob und inwiefern übergeordnete Ziele gefährdet sein können und damit volkswirtschaftliche Wohlfahrtsverluste verbunden sind. Für die Finanzregulierung sind dies primär die Ziele der Finanzstabilität, der Finanzmarktintegrität und des (kollektiven) Verbraucherschutzes.⁷

Evidenz kann dabei verschiedene Ausprägungsformen haben. Etwa die der Beobachtung von Trends und Verhaltensweisen, welche möglicherweise auf ein Problem hinweisen. Idealerweise wird diese Evidenz dann mit weiteren Daten und einer systematischeren (empirischen) Analyse unterlegt. Auch sollte die Größenordnung und Relevanz eines Problems identifiziert werden.

Nunmehr sollte in einem Impact Assessment evidenzbasiert eruiert werden, ob ein Markt- oder Regulierungsversagen vorliegt, das ein regulatorisches Eingreifen notwendig macht. Gründe für Marktversagen sind asymmetrische Informationsverteilung, ein Mangel an Wettbewerb, negative externe Effekte oder die mangelnde Bereitschaft, zu öffentlichen Gütern wie der Finanzstabilität beizutragen.

Ohne eine entsprechende Evidenz ist hierbei weder die Identifikation von Markt- oder Regulierungsversagen möglich noch ist es möglich, die richtige regulatorische Antwort in Form von alternativen (zu bewertenden) regulatorischen Maßnahmen zu finden. Dabei sind die Anreize

5 Das Lamfalussy-Verfahren unterteilt den Europäischen Finanzregulierungsprozess in mehrere Stufen. Weitere Infos finden sich hier: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-reforms-and-their-progress/regulatory-process-financial-services/regulatory-process-financial-services_en

6 Die Guidelines der 3L3 Vorgängerbehörden sind inhaltlich noch als Maßstab zu sehen und werden auch noch aktiv als Arbeitsgrundlage genutzt.

7 Vgl. u.a. Finanzdienstleistungsaufsichtsgesetz vom 22. April 2002 (BGBl. I S. 1310), das zuletzt durch Artikel 14 Absatz 3 des Gesetzes vom 17. Juli 2017 (BGBl. I S. 2446) geändert worden ist.

für die Marktteilnehmer zu berücksichtigen und die Auswirkungen der regulatorischen Maßnahmen sowohl in positiver als auch negativer Hinsicht zu ermitteln. Bei dieser Kosten-Nutzen-Analyse sind daher unter der Rubrik Kosten nicht lediglich Bürokratiekosten, sondern auch unerwünschte Nebeneffekte und mithin volkswirtschaftliche Wohlfahrtsverluste zu berücksichtigen.

Sofern Regulierungsvorhaben bereits in ähnlicher Form angewendet oder auch in anderen Jurisdiktionen genutzt wurden, kann für die Folgenabschätzung empirische Evidenz verfügbar sein. Die Evidenz für die Folgenabschätzung ist jedoch nicht notwendigerweise ausschließlich empirischer Natur. Liegen solche Erfahrungswerte nicht vor, muss auf modell- und experimentbasierte Evidenz zurückgegriffen werden. Ökonomische Modelle und experimentelle Studien sind zudem immer hilfreich, um mögliche Anreize, die das Verhalten der Marktteilnehmer beeinflussen können, zu verstehen.

Die Konsultation von Marktteilnehmern, der Öffentlichkeit, aber auch der Wissenschaft spielt bei der Generierung von Evidenz daher eine maßgebliche Rolle.

Aus Sicht von Institutionen der Finanzmarktaufsicht und -regulierung wie der BaFin sind hierbei auch die Umsetzbarkeit und der Aufwand von Regulierung und Aufsicht von zentraler Bedeutung.

Der Kreis schließt sich, wenn regulatorische Maßnahmen im Nachhinein evaluiert werden (Review-Prozess), um zu ermitteln, ob die gewünschten Effekte erzielt wurden oder unerwünschte Nebeneffekte im Sinne eines Regulierungsversagens eingetreten sind. Auch hier ist Evidenz von zentraler Bedeutung, da sie für die Regulierung eine informationsbasierte Entscheidungsgrundlage schafft.

Naturgemäß findet der regulatorische Prozess nicht im luftleeren Raum statt. Mithin kann sich ein Spannungsfeld zwischen politischen Erwägungen und evidenzbasierter Regulierung ergeben. Die evidenzgestützte Analyse schafft jedoch die Voraussetzungen dafür, Sachverhalte möglichst objektiv zu betrachten und auf Seiten der Politik das Verständnis für regulatorische Maßnahmen zu fördern. Ziel muss sein, das im Sinne möglichst geringer Wohlfahrtsverluste bestmögliche Ergebnis zu erzielen. Dies bedeutet keineswegs, dass ein rein technokratisches Vorgehen richtig oder möglich wäre. Vielmehr müssen auch im Rahmen der evidenzbasierten Analyse und der Folgenabschätzung regulatori-

scher Maßnahmen politische Erwägungen berücksichtigt werden. Dies wird insbesondere dann deutlich, wenn auch die übergeordneten Ziele in Spannung zueinanderstehen. Beispielsweise kann die Sicherung der Solvenz von beaufsichtigten Banken oder Versicherungsunternehmen aus Finanzstabilitätsgründen im Konflikt mit dem Verbraucherschutz stehen. Zudem gibt auch die beste Evidenz niemals vollständige Sicherheit, was die Folgen eines regulatorischen Eingreifens angeht. Der Review-Prozess darf daher nicht als notwendiges Übel, sondern als Notwendigkeit verstanden werden. Auch kann meist erst bei einem solchen Review ein Marktversagen objektiv von einem Regulierungsversagen abgegrenzt werden.

Um die richtige Balance zwischen evidenzbasierter Erkenntnis und politischen Erwägungen zu finden, ist der Austausch zwischen Institutionen der Regulierung und Aufsicht mit der Politik maßgeblich. Die Bundesanstalt für Finanzdienstleistungsaufsicht (BaFin) steht im regelmäßigen Austausch mit den relevanten Schnittstellen zu den Entscheidungsträgerinnen und Entscheidungsträgern in der Politik, insbesondere dem Bundesministerium der Finanzen, sowie europäischen Gremien, um die verschiedenen Phasen des Regulierungsprozesses zu begleiten und ihre fachliche Expertise in die politischen Erwägungen einzubringen. Dabei generiert und stellt die BaFin im Rahmen ihrer Zuständigkeit in umfangreichem Maß empirische Evidenz zur Verfügung.

Zum Beispiel wird die BaFin im Rahmen von RIA-Analysen auf EU-Ebene konsultiert und steuert ihre eigene Expertise und Evidenz bei und – mittelbar – die der Aufsichtsobjekte, an die sie Umfragen und Auskunftersuchen richtet. Durch diese mittelbare Konsultation der Industrie können deren Aussagen strukturiert und relativiert werden – etwa dann, wenn Lobbyismus-Erwägungen die Aussagen verzerren.

Auf EU-Ebene erscheint bei der Evaluierung von Regulierung vor allem eine Einbindung der Expertinnen und Experten bereits auf Level 1 sinnvoll, da auf Level 2 und 3 meist nur noch Ausgestaltungsmerkmale von regulatorischen Maßnahmen evaluiert werden.

Erst eine evidenzbasierte Analyse ermöglicht eine rationale Regulierung, die unter Abwägung der verschiedenen Interessen bestmögliche Ergebnisse für die Wohlfahrt im Rahmen eines demokratischen Prozesses produziert, nämlich anhand der Antwort auf die Frage: „Welches wäre die richtige Regulierungsmaßnahme und wie setze ich sie um?“

Ein Bericht zur Folgenabschätzung (*impact assessment report*) darf jedoch nicht so verfasst sein, dass er die politische Entscheidung für eine Regulierungsmaßnahme automatisch rechtfertigt. Es sollte vermieden werden, dass eine möglicherweise ex ante präferierte Entscheidung bzw. Option als „alternativlos“ betrachtet wird. Eine kritische Auseinandersetzung mit den Gesetzesfolgen muss sichergestellt werden. Entscheidend ist u.a., welche Frage gestellt wird. Falsch wäre die Frage: „Wie setze ich die gewünschte Regulierung um?“ Richtig wäre beispielsweise die Frage: „Welches wäre die richtige regulatorische Maßnahme, um einem Marktversagen zu begegnen?“ Grundsätzlich wird die Einschätzung in der Praxis häufig dadurch erschwert, dass positiver Wohlfahrtsnutzen und unerwartete Nebeneffekte schwerer zu messen sind als reine Bürokratiekosten.

Unabhängig vom gesetzgeberischen und regulatorischen Prozess auf europäischer oder nationaler Ebene im engeren Sinne findet auch bei der Risikoanalyse durch Aufsichts- und Regulierungsgremien eine evidenzbasierte Analyse und Informationsfindung statt. Insbesondere dann, wenn eine solche Analyse präventiv Risiken ermitteln soll, kann Evidenz nur unter Unsicherheit und auf Basis von Annahmen gewonnen werden. Dies kann die Ableitung möglicher Implikationen erheblich erschweren.

Das Financial Stability Board (FSB), die globalen Standardsetzer, das European Systemic Risk Board (ESRB) und andere makroprudenzielle Institutionen, die ESAs, nationale Aufsichtsbehörden und Zentralbanken versuchen, Risiken für die Finanzstabilität zu identifizieren. Dabei spielen die Evidenz und der Austausch mit Marktteilnehmern und der Wissenschaft eine wichtige Rolle. Eine Möglichkeit, diesen Austausch zu institutionalisieren, sind regelmäßige Foren sowie wissenschaftliche Beiräte bzw. Gremien und Auftragsstudien, wie sie auch in der Politikberatung eingesetzt werden.

Beispiele hierfür sind der Fachbeirat der BaFin, der wissenschaftliche Beirat des ESRB (Advisory Scientific Committee – ASC) und die Gruppe der ökonomischen Berater (Group of Economic Advisors – GEA), die das Marktanalyse-Komitee (Committee of Economic Market Analysis – CEMA) der European Securities and Markets Authority (ESMA) berät.

Anhand von definierten Themen und Fragen soll in diesen Gremien die akademische Kompetenz genutzt werden, Erkenntnisse und – im

weiteren Sinne – Evidenz für Probleme und aufsichtsbezogene bzw. regulatorische Fragen zu gewinnen. Wichtig bei der Einrichtung solcher Gremien ist, dass diese lediglich der Unterstützung der Entscheidungstragenden in den relevanten Gremien dienen. Ihre Aufgabe ist nur, eine informationsbasierte Entscheidungsgrundlage aus Evidenzsicht zu generieren. Das Agenda-Setting und die finale Entscheidung, welche die oben erwähnte Balance mit politischen Erwägungen schafft, müssen bei den hierfür demokratisch legitimierten Entscheidungstragenden verbleiben. Keinesfalls darf hieraus ein akademisch-technokratischer Automatismus entstehen.

Unabhängig davon kann es auch Interessenkonflikte mit Wissenschaftlerinnen und Wissenschaftlern sowie administrative Hürden geben. Für Wissenschaftlerinnen und Wissenschaftler liegt ein wesentlicher Nutzen eines Engagements in einem solchen Gremium und damit die Bereitschaft, an einem Projekt mitzuwirken, in einem exklusiven Datenzugang und somit der Möglichkeit, einzigartige Veröffentlichungen verfassen zu können. Die Daten, über die eine Finanzaufsicht verfügt, sind jedoch zumeist vertraulich und können für Forschungszwecke nicht oder allenfalls anonymisiert verwendet werden. Es ergibt sich mithin das Problem, Wissenschaftlerinnen und Wissenschaftlern einen sicheren Zugang zu vertraulichen Daten zu schaffen. Die Lösung ist juristisch und praktisch nicht immer einfach. So müssen bspw. entsprechende Vertraulichkeitsvereinbarungen getroffen und Datenschutzräume eingerichtet werden. Oft ist der Zugang gar nicht möglich, da die Daten lediglich für Zwecke der Aufsicht verwendet werden dürfen und sich juristisch die Frage stellt, welche ökonomischen Fragestellungen hierunter subsumiert werden können.

Eine praktische Möglichkeit könnte sein, dass Doktorandinnen und Doktoranden befristet bei einer Behörde eingestellt werden. Damit unterlägen sie dann den Datenschutzvorschriften für Beschäftigte der Behörde und dennoch wäre der Transfer von Expertise ermöglicht. Auch die finanzielle Entschädigung und Entlohnung externer Expertinnen und Experten wäre zu klären.

Fazit

Evidenzbasierte Finanzmarktregulierung ist wichtig, um einen objektivierten Entscheidungsprozess auf einer guten Informationsgrundlage

zu schaffen. Evidenz leistet einen maßgeblichen Beitrag zum Regulierungsprozess und demokratisch legitimierte Entscheidungsträgerinnen und -träger sollten sie bei politischen Erwägungen berücksichtigen. Rein technokratisches Vorgehen einerseits und politisch motivierte Voreingenommenheit andererseits sollten vermieden werden. Es gilt, das bestmögliche Ergebnis unter dem Gesichtspunkt der gesamtwirtschaftlichen ökonomischen Wohlfahrt zu finden.

Literatur

Europäische Kommission (o.J.). „Bessere Rechtsetzung: warum und wie?“

https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/better-regulation-why-and-how_de [Zugriff Juli 2018]

Committee of European Banking Supervisors, Committee of European Securities Regulators, Committee of European Insurance and Occupational Pensions Supervisors (2008). „Impact Assessment Guidelines for EU Lamfalussy Level 3 Committees“, Ref: CESR/07-089b , Ref: CEBS 2007 58 , Ref: CEIOPS-3L3-08/06

<https://www.eba.europa.eu/documents/10180/105244/3L3.IA.GL.pdf> [Zugriff Juli 2018]

Evidenzbasierte Evaluierung als Voraussetzung für Reformen der Finanzmärkte – Ursachen und Schlussfolgerungen aus der Finanzkrise 2008

Von JÖRG ROCHOLL

Korrespondenzadresse:

ESMT Berlin

Schloßplatz 1

10178 Berlin

joerg.rocholl@esmt.org

Politische und regulatorische Entscheidungen basieren auf Fakten, so die natürliche Annahme. Auch bei der Reform von Finanzmärkten werden Entschlüsse getroffen, weil sich die Entscheidungsträgerinnen und -träger hinreichend mit dem Für und Wider, den Kosten, Nutzen und gegebenen Konsequenzen auseinandergesetzt haben. Auf Basis der gesammelten Daten und nach der Analyse aller beteiligten Elemente wird folglich eine möglichst objektive Entscheidung getroffen, so die Annahme.

In der Realität spielen jedoch häufig verschiedene andere Faktoren eine ernstzunehmende Rolle: Wer hat eine Aussage getroffen (Eminenz)? Mit wie viel Nachdruck hat er seinen Standpunkt dargelegt (Vehemenz)? Wie elegant hat er die Aussage formuliert (Eloquenz)? Und welche finanziellen Mittel stehen ihm zur Verfügung, um sein Ziel zu verfolgen? Letztere sind insbesondere dann kritisch, wenn der politische Druck hoch ist. Dann entscheiden die bessere Lobby und umfangreichere finanzielle Mittel darüber, was letzten Endes als gültige Meinung akzeptiert wird.

Um für alle Beteiligten sinnvolle und langfristig wirksame Entscheidungen zu treffen, ist jedoch, gerade für Finanzmärkte und ihre Reform, eine evidenzbasierte Evaluierung unabdingbar. Ein besonders offensichtlicher Beleg dafür ist die Frage nach den Ursachen der Finanzkrise im Jahr 2008 – und den richtigen Schlussfolgerungen daraus.

1. Wer waren die Hauptverantwortlichen der Finanzkrise vor zehn Jahren?

Die Benennung eines einzigen Hauptverantwortlichen gestaltet sich schwierig. Vielmehr werden — in Abhängigkeit von (Teil-)Evidenz

und Interessenlage der Berichterstatterinnen und Berichterstatter — verschiedene Akteure genannt. Zu ihnen gehören:

- Banken, aufgrund ihrer fragwürdigen und übermäßig riskanten Geschäftsmodelle und -praktiken
- Zentralbanken, gerade in den USA, aufgrund ihrer zu expansiven Geldpolitik im Vorfeld der Finanzkrise
- Regierungen, zum Beispiel durch die Begünstigung unpräziser Kreditvergabestandards in den USA, die fehlende Konsolidierung des Bankensektors in Europa oder eine übermäßige öffentliche Verschuldung
- Regulierungsbehörden, aufgrund der Festlegung unzureichender Kapital- und Liquiditätsanforderungen
- Bankenaufsichten, aufgrund ihrer Distanz bzw. Distanzlosigkeit gegenüber Banken
- Ratingagenturen, aufgrund der Vergabe fragwürdiger und zu optimistischer Ratings
- Konsumenten und Konsumentinnen, aufgrund übermäßiger Kreditaufnahme und überbordender privater Verschuldung.

Das verantwortungslose Verhalten dieser und anderer Akteure ab 2007 hatte massive Verluste zur Folge: Nicht nur in Deutschland reduzierte sich das Wirtschaftswachstum anhaltend. Im Jahre 2009 verzeichnete die Bundesrepublik sogar ein negatives Wachstum (BIP-Wachstumsrate -5,0%¹). Die Steuerzahlenden wurden massiv belastet und das Vertrauen in die (soziale) Marktwirtschaft erschüttert. Schnell wurden deshalb Forderungen nach einer Verbesserung der regulatorischen Rahmenbedingungen laut.

2. Regulatorische Reaktionen, mit besonderem Fokus auf die Banken

Eine Reihe eingeführter regulatorischer Reaktionen versprach Besserung: Die Kapitalanforderungen, sowohl risikogewichtet als auch absolut, wurden erhöht. Der Basler Ausschuss entwickelte Liquiditätsanforderungen auf internationaler Ebene, z. B. die Liquidity Coverage Ratio (LCR) und

1 STATISTISCHES BUNDESAMT (Hrsg.) 2010.

die Net Stable Funding Ratio (NSFR). Im Rahmen des Single Supervisory Mechanism (SSM) wurde die Bankenaufsicht in Europa vereinheitlicht. Die Einführung des Single Resolution Mechanism (SRM) verbesserte und vereinheitlichte die Abwicklung gefährdeter Banken. Zusätzlich wurden die Einlagensicherungssysteme harmonisiert, ohne jedoch vergemeinschaftet zu werden.

Seit der Finanzkrise und der Einführung konsekutiver regulatorischer Mechanismen haben sich zwei Lager herausgebildet. Die Anhängerinnen und Anhänger des einen sprechen sich für eine Hochhaltung des Regulierungstempos aus. Nur so kann man ihres Erachtens sichergehen, dass sich die Ereignisse des Jahres 2008 und vor allem die massiven Verluste für den Steuerzahler nicht wiederholen. Anhängerinnen und Anhänger des Gegenlagers sprechen von einem „Regulierungstsunami“ und fordern ein Sabbatical, eine Latenzphase, in der vorerst keine weiteren regulatorischen Maßnahmen eingeführt werden. Sie verweisen auf die vermeintlich weitreichenden negativen Auswirkungen der Regulierungsschritte auf die Kreditvergabe an Unternehmen und Privatkonten. Zudem bemängeln sie die fehlende Risikogewichtung der Leverage Ratio.

Fazit: Über die Auswirkungen der bisherigen Reformen, ihre Kosten (wie den bürokratischen Aufwand bei Informations- und Meldeanforderungen), ihren Nutzen und die möglichen Wechselwirkungen zwischen den verschiedenen Regulierungsschritten ist zu wenig bekannt. Der Fokus der Debatte liegt noch immer auf einzelnen Maßnahmen und Sektoren anstelle einer auf Konsistenz gerichteten ganzheitlichen Betrachtung.² Gleichzeitig lässt sich der Eindruck eines gewissen Aktionismus nicht von der Hand weisen: Ein gutes Beispiel hierfür sind die eilig beschlossenen und international zunächst nicht abgestimmten Regelungen zu Trennbanken vor der Bundestagswahl im Jahr 2013.

3. Was ist zu tun?

Trotz verschiedener Ansätze fehlen bisher großflächige und glaubwürdige Studien zu den Auswirkungen der eingeführten Regulierungsmechanismen. Auch Kosten und Nutzen der Maßnahmen müssen auf deutscher, europäischer und globaler Ebene untersucht werden, bevor weitere sinnvolle Schritte eingeleitet werden können.

2 DEUTSCHE BUNDESBANK 2012.

Die Durchführung methodologisch belastbarer und repräsentativer Studien in Form von Kosten-Nutzen-Analysen, Auswirkungsstudien, akademischer Forschung und Programmevaluierung ist kein Sabbatical im Sinne eines Nichtstuns. Vielmehr dienen diese Studien als wichtige und sinnvolle Nutzung der Zeit, um angemessene nächste Schritte zu definieren. Die Ergebnisse könnten helfen, die Diskussion zu versachlichen und sie auf eine stabile Basis allgemein akzeptierter Fakten zu stellen.

4. Was sind die wesentlichen Voraussetzungen?

- I) Zunächst müssen die relevanten Fragestellungen identifiziert und so eine realistische Erwartungshaltung geschaffen werden. Viele relevante Fragen werden nur langsam und möglicherweise eingeschränkt empirisch zu beantworten sein. Es gilt, ex ante ein gemeinsames Verständnis über die zu beantwortenden Fragen und ihre Relevanz zu schaffen.³ Die politische Unterstützung des G20-Gipfels in Hamburg 2017 lässt hier hoffen: Die Ergebnisse der laufenden Projekte, allen voran zu den OTC-Derivaten und der Finanzierung der Infrastruktur, werden beim nächsten G20-Gipfel in Argentinien im Herbst 2019 präsentiert.
- II) Eine zweite Grundlage – gerade in Zeiten von Big Data und immer höheren Verarbeitungsgeschwindigkeiten – ist die Bereitstellung der wichtigsten Daten für alle Akteure. Hier gibt es vielversprechende Ansätze, zum Beispiel das Forschungsdaten- und Servicezentrum der Bundesbank (FDSZ). Auch die Vernetzung mit anderen nationalen Datenbanken, zum Beispiel der Bundesagentur für Arbeit, ist aussichtsreich. Langfristig könnte so beispielsweise die Frage beantwortet werden, inwiefern sich ein vermindertes Kreditangebot auf die Arbeitsnachfrage auswirkt. Auch die engere internationale Zusammenarbeit verschiedener Zentralbanken und anderer internationaler Organisationen ist diesbezüglich zukunfts-trächtig.

3 FINANCIAL STABILITY BOARD 2017.

- III) Des Weiteren stellt sich die Frage nach der Kausalität exogener Schocks im Allgemeinen und der Finanzkrise im Besonderen. Welche Voraussetzungen waren gegeben? Welche Ursachen kann man in Zukunft beeinflussen? Welche Warnsignale sind schon früh erkennbar? Die Analyse des Kontrafaktischen steht hierbei im Mittelpunkt: Was wäre passiert, wenn eine bestimmte Entscheidung nicht getroffen worden wäre? Im Idealfall sollte es hierzu isolierte natürliche Experimente geben. De facto aber existieren solche Experimente nicht im dafür nötigen Umfang. Zur Illustration der damit verbundenen Herausforderungen sollen drei Beispiele dienen:

Erstes Beispiel

In der Geldpolitik stellt sich die Frage, ob die Quantitative Lockerung (QE) geholfen hat, die Inflationsrate in Europa an das gewünschte Niveau heranzuführen. Man kann in beide Richtungen argumentieren: Einerseits lag die Inflationsrate nach Beginn von QE immer noch unterhalb des angestrebten Wertes. Andererseits hätte sie ohne eine quantitative Lockerung noch deutlich niedriger liegen können. Zudem ist nicht nur die Richtung der Wirkung entscheidend, sondern auch ihr quantitativer Effekt, d.h. die Elastizitäten, die eine Regulierungsmaßnahme bewirkt.

Zweites Beispiel

Als Folge der Finanzkrise wurde eine Reihe von Regulierungsmaßnahmen simultan eingeführt. Doch welche dieser Maßnahmen war die entscheidende? Man könnte behaupten, dass die Einführung der Bankenunion im Jahre 2012 die Stabilität und Überlebenschancen des Euro erhöht hat. In jenem Jahr sagte der Präsident der Europäischen Zentralbank Mario Draghi jedoch in seiner vielzitierten Rede, er würde tun, „whatever it takes“⁴, um den Euro zu erhalten. Welcher Mechanismus hat also wie viel Einfluss? Wichtig ist es, langfristig die Möglichkeit zu schaffen, einzelne Maßnahmen von anderen Maßnahmen in ihrer Wirkung isolieren zu können.

4 EUROPEAN CENTRAL BANK 2012.

Drittes Beispiel

Zwei identische Hotels in Tirol und Südtirol nehmen Kredite zur Finanzierung auf. Eines der Hotels liegt in Österreich, das andere in Italien. Zusätzliche Faktoren wie Wetter und Lage stellen keinen Unterschied dar. Das italienische Hotel zahlt höhere Kreditzinsen als das österreichische.

Man kann nun behaupten, dass der Unterschied durch die unterschiedlichen Risiken der jeweiligen Staaten begründet ist, die eine Auswirkung auf die Kreditzinsen und das Denominationsrisiko haben. Unter dieser Annahme stellt eine expansive EZB-Geldpolitik zur Schaffung eines besseren Finanzierungszugangs möglicherweise einen effektiven Lösungsansatz dar. Andererseits könnte man argumentieren, das Insolvenzregime in Österreich sei deutlich effizienter und damit gläubigerfreundlicher als in Italien. Unter diesen Bedingungen wären die geringeren Zinssätze in Österreich gerechtfertigt und eine Verbesserung des Insolvenzregimes in Italien folglich die richtige politische Antwort.

- IV) Die vierte Voraussetzung für den erfolgreichen Einsatz evidenzbasierter Evaluation ist die Sicherstellung der Reproduzierbarkeit der Ergebnisse. Sie sollten verallgemeinerbar und aggregierbar sein, damit sie als kontinuierlicher Bestandteil des wissenschaftlichen Diskurses immer wieder begutachtet und neu interpretiert werden können. Derzeit gibt es jedoch kaum Anreize für Wissenschaftlerinnen und Wissenschaftler, auf die Reproduzierbarkeit bestehender Studien zu achten. Auch eine Würdigung solcher Studien im akademischen Sektor bleibt bislang aus. Entsprechend fehlt es an belastbaren und vergleichbaren Datensätzen, die in der Finanzpolitik genutzt werden können.

Gelten die vorhandenen Studien also nur für bestimmte Länder, Zeiträume oder unter bestimmten institutionellen Voraussetzungen? Die mangelnde Reproduzierbarkeit der Studien veranschaulicht eine Leerstelle im System: Die Einführung von Clearingstellen könnte helfen, in Zukunft bestehende Ergebnisse sammeln, systematisieren und aggregieren zu können. Ziel soll es sein, ein Grundverständnis darüber zu schaffen, was wir wissen und was nicht. Gleichsam bedarf es der Erstellung von Metastudien, solcher Studien also, die die be-

reits existierenden Forschungsergebnisse miteinander vergleichen und als Kollektiv auswerten. Mittels unterschiedlicher empirischer Designs können dann beispielsweise folgende Fragen beantwortet werden: Was wissen wir zu einem bestimmten Thema? Worauf können wir uns einigen?

- V) Schließlich müssen die gesammelten Daten in die Praxis übertragen werden. Die Ergebnisse sollten in der Politik anwendbar und regulierbar sein. Für den erfolgreichen Transfer von der Theorie in die Praxis bedarf es eines engen Austausches und verbesserter Kommunikation zwischen Wirtschaft, Wissenschaft und Politik. Mitarbeiterinnen und Mitarbeiter sollten bereits in Aus- und Weiterbildung über die zentralen Aspekte wie Kausalität und Endogenität sowie Robustheit und Aggregierbarkeit der Daten informiert werden. Nur so kann ein gemeinsames Verständnis der zu bearbeitenden Fragen entwickelt werden. Ob sich die angewandten Methoden eignen, um in der Praxis anwendbare Ergebnisse zu ermitteln und welche Aussagekraft diese Ergebnisse für die Entwicklungen der Finanzpolitik haben, wird sich erst auf lange Sicht gesehen beurteilen lassen.

5. Ausblick

Die evidenzbasierte Politikberatung ist von essentieller Bedeutung und hat großes Potential. Neben den oben genannten Herausforderungen warten weitere wichtige Fragen auf eine angemessene Beantwortung. Wie gut ist beispielsweise der Finanzierungszugang kleiner und mittelständischer Unternehmen (KMU)? Welche langfristigen Auswirkungen hat die Vereinheitlichung der Aufsicht über Banken und andere systemrelevante Finanzinstitute („too big to fail“) im Rahmen des Single Supervisory Mechanism (SSM)? Forschungsarbeiten zu diesen und anderen Themen werden beim Financial Stability Board (FSB) im Laufe des Jahres aufgenommen.

Die Grundlagen der Erkenntnisse aus verschiedenen Disziplinen müssen miteinander abgeglichen und daraus für die Praxis relevante Erkenntnisse für die zukünftige Anwendung abgeleitet werden. Der Einsatz evidenzbasierter Evaluation als Basis für Entscheidungen in der Finanzpolitik hat gute Zukunftsaussichten. Es bedarf jedoch einer erhöhten Kommunikation unter den verschiedenen Akteuren und der Erfüllung der

hier genannten Voraussetzungen. Die Implementierung evidenzbasierter Evaluation im Bereich der Finanzmärkte muss in jedem Fall langfristig gedacht werden.

Literatur

Jayaraman, R., Rocholl, J. (2017). „Researched-Based Policy Advice to the G20“.

http://www.g20-insights.org/wp-content/uploads/2017/05/OV_Re-search-based-policy-advice.pdf [Zugriff 12. Juni 2018]

Leuz, C. (2018). „Evidence-Based Policymaking: Promise, Challenges and Opportunities for Accounting and Financial Markets Research“. European Corporate Governance Institute (ECGI) – Law Working paper 397.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3161401 [Zugriff 12. Juni 2018]

Statistisches Bundesamt (Hrsg.) (2010). „Gesamtwirtschaftliche Konjunkturdaten für Deutschland, Bruttoinlandsprodukt 2009 für Deutschland - Begleitmaterial zur Pressekonferenz am 13. Januar 2010 in Wiesbaden“.

https://www.destatis.de/DE/PresseService/Presse/Pressekonferenzen/2010/BIP2009/Pressebrochure_BIP2009.pdf [Zugriff 05. Juli 2018]

Deutsche Bundesbank (2012). „Geschäftsbericht“.

https://www.bundesbank.de/Redaktion/DE/Downloads/Veroeffentlichungen/Geschaeftsberichte/2012_geschaeftsbericht.pdf?__blob=publicationFile [Zugriff 12. Juni 2018]

Financial Stability Board (2017). „Proposed Framework for Post-Implementation Evaluation of the Effects of the G20 Financial Regulatory Reforms“.

<http://www.fsb.org/wp-content/uploads/Framework-for-the-post-implementation-evaluation-of-the-G20-financial-regulatory-reforms.pdf> [Zugriff 12. Juni 2018]

European Central Bank (2012). „Verbatim of the Remarks Made by Mario Draghi. Speech by Mario Draghi, President of the European Central Bank, at the Global Investment Conference in London 26 July 2012“. <https://www.ecb.europa.eu/press/key/date/2012/html/sp120726.en.html> [Zugriff 14. Juni 2018]

Gute Evidenz für wirksame Fiskalpolitik

Von WERNER EBERT, ELMAR DÖNNEBRINK und LUDGER SCHUKNECHT¹

Korrespondenzadresse:

Bundesministerium der Finanzen

Wilhelmstraße 97

10117 Berlin

Postanschrift: 11016 Berlin

Werner.Ebert@bmf.bund.de

1. Einführung

Eine Lehre aus der Finanz-, Wirtschafts- und Staatsschuldenkrise ist die fundamentale Bedeutung einer vorausschauenden Fiskalpolitik für das effiziente und nachhaltige staatliche Handeln. Diese Erkenntnis hat in eine Vielzahl von Studien internationaler Institutionen zum Mapping fiskalischer Risiken und zu fiskalischen Regeln der Haushaltspolitik Eingang gefunden. Der Bedarf nach einer guten Risikoanalyse spiegelt sich in Konzepten wie Fiscal Heat Maps (IWF)² oder dem Debt Sustainability Monitor (EU-Kommission)³ und nicht zuletzt in der makro- und mikroprudenziellen Überwachung des Finanzmarktbereichs wider. Zudem unterstreichen sie die Bedeutung institutioneller Absicherungsmechanismen wie des Stabilitäts- und Wachstumspakts in Europa oder der Schuldenbremse in Deutschland.

Auch und gerade in wirtschaftlich guten Zeiten gilt es, die Sensibilität für fiskalische – exogene wie auch systemimmanente – Ineffizienzen und Risiken wach zu halten und sich durch richtige Prioritätensetzungen für die Herausforderungen der Zukunft zu wappnen. Der Handlungsdruck besteht fort; nicht nur um kurzfristige Krisen zu bewältigen, sondern auch um gravierende langfristige Herausforderungen⁴ rechtzeitig anzugehen, die ansonsten größere ökonomische, gesellschaftliche und

1 Kontakt Ludger Schuknecht: Ludger.Schuknecht@OECD.org

Die Autoren vertreten in diesem Beitrag ausschließlich ihre persönliche Auffassung und nicht notwendigerweise die des BMF.

2 Siehe aktuell IMF 2018.

3 EU-KOMMISSION 2018.

4 Zu Herausforderungen der Altersversorgung beispielsweise BERENBERG BANK 2018.

nicht zuletzt fiskalische Verwerfungen auslösen können.

Dabei ist das übergreifende finanzpolitische Ziel in Deutschland klar definiert: Der strukturell weitgehend ausgeglichene Bundeshaushalt als verfassungsrechtliche Bindung sowie die politische Zielsetzung eines ohne Nettoneuverschuldung ausgeglichenen Bundeshaushalts („schwarze Null“) bilden den finanzpolitischen Rahmen, innerhalb dessen Prioritäten gesetzt und Vorsorge für zukünftige Risiken getroffen werden. Das Zielsystem auf der Makroebene ist transparent und erlaubt eine einfache Ex-post-Kontrolle durch jedermann.

Zu den Abwägungen über die richtigen Prioritäten der Politik gehört, die öffentlichen Ausgaben und Einnahmen unter dem Blickwinkel der Effektivität und Effizienz und im Hinblick auf ihre Zielerreichung zu prüfen, aber auch die Ziele selbst zu hinterfragen. Nebenbei bemerkt erscheint die bisweilen nahezu vergessene deutsche Tradition der Ordnungspolitik, gemessen an diesen Zielstellungen, keineswegs altmodisch und kann als politisch-ökonomischer Kompass dienen.

2. Wirkungsorientierung des Budgets als Fokus

Für den Bund haben die Schuldenbremse und das zur haushaltspolitischen Handhabung derselben im Jahr 2012 eingeführte Top-down-Verfahren (Haushaltseckwerteverfahren) einen neuen institutionellen Rahmen aufgespannt. Anders als früher müssen die Ressorts Haushaltsansätze nicht mehr im Rahmen der Haushaltsaufstellung im Einzelnen mit dem Bundesfinanzministerium verhandeln, sondern übernehmen stärker Eigenverantwortung für die Ressourcenallokation. Auf Vorschlag des Bundesfinanzministeriums beschließt das Kabinett im Top-down-Verfahren das Gesamtbudget und dessen Aufteilung auf die Ressorts. Die Fachressorts planen die Aufteilung ihrer Haushaltsmittel weitgehend autonom.⁵

Vor Einführung des Top-down-Verfahrens lag die Verantwortung für den Haushaltsausgleich im Haushaltsaufstellungsprozess beim Bundesfinanzministerium, mit dem jeder von den Ressorts angemeldete Bedarf einzeln zu verhandeln war. Das Finanzministerium musste die einzelnen Bedarfe der Ressorts untereinander und in der Gesamthöhe austarieren und den insgesamt vorhandenen Finanzierungsspielräu-

5 Vgl. BUNDESFINANZMINISTERIUM 2017a.

men anpassen. Das Haushaltsvolumen des Regierungsentwurfs war Ergebnis dieses Verhandlungsprozesses.

Mit den durch das Top-down-Verfahren neu gewonnenen Freiheiten der Ressorts stellt sich die Frage, wie eine Haushaltssteuerung erreicht werden kann, mit der mehr „value for money“ für Bevölkerung und Wirtschaft gehoben werden kann. Denn einerseits ermöglicht die dezentrale Ressourcenverantwortung die Entscheidung über die Ressourcenallokation im jeweiligen Politikfeld auf einer guten Informationsgrundlage, andererseits muss der Mitteleinsatz in den Politikfeldern weiterhin koordiniert werden und ein effizientes System von „Checks und Balances“ installiert sein.

Aufgrund der kontinuierlich guten Entwicklung der öffentlichen Finanzen im letzten Jahrzehnt steht zudem die Feuerprobe aus, wie Ausgabenkürzungen im Top-down-Verfahren umgesetzt werden könnten. Einerseits ist naheliegend, dass solche Kürzungen im Rahmen verhandelter Eckwerte in Eigenverantwortung der Ressorts vorgenommen werden, andererseits stellt sich die Frage, ob ressortübergreifende Wirksamkeitsanalysen helfen können, notwendige Prioritätensetzungen und sich daraus ergebende Anpassungen vorzunehmen. Allerdings besteht auch in „good times“ die Gefahr, dass investive Ausgaben verdrängt werden und damit die Qualität des öffentlichen Haushalts erodiert.⁶

Vor diesem Hintergrund hält der Koalitionsvertrag der neuen Großen Koalition fest, dass „durch umfassende Aufgabenkritik sowie durch die regelmäßige Überprüfung der Maßnahmen auf Effektivität und Effizienz zusätzliche Spielräume“ erarbeitet werden sollen. Dazu sollen die Ansätze zur Verbesserung der Wirkungsorientierung des Haushalts, zum Beispiel einnahme- und ausgabeseitige Haushaltsanalysen, gestärkt und weiterentwickelt werden.

Der Wirkungsorientierung des Haushalts wird damit eine wichtige Funktion zugeschrieben, nicht zuletzt, um Spielräume zu erschließen, die für Umschichtungen innerhalb des Etats genutzt werden können. Das Instrument der Spending Reviews ist ein wesentlicher Baustein zur Verbesserung der angestrebten Wirkungsorientierung. Andere Instru-

6 Vgl. bereits früh AFONSO, EBERT, SCHUKNECHT, THÖNE 2004 oder SCHUKNECHT, ZEMANEK 2018.

mente wie etwa Makroanalysen der Qualität der öffentlichen Finanzen, Evaluierungen, ein an Leitlinien ausgerichteter Subventionsbericht oder die Nutzung marktorientierter Instrumente der Bereitstellung öffentlicher Leistungen, wie etwa öffentlich-private Partnerschaften, ergänzen das Instrumentarium.

Innerhalb der Bundesregierung hat das Bundesfinanzministerium schon frühzeitig auf die Notwendigkeit eines erweiterten Instrumentariums zur Koordinierung und Wirkungsorientierung des Haushalts hingewiesen und dessen Entwicklung nachdrücklich unterstützt.

Um diese Aufgaben erfüllen zu können, muss Politik über die notwendigen Informationen verfügen, die Risiken kennen und Abstimmungsmechanismen für Politikverbesserungen entwickeln. Im Folgenden werden drei Kernelemente der regel- und evidenzbasierten Finanzpolitik vorgestellt: die Tragfähigkeitsanalyse, die Spending Reviews und die Subventionsberichterstattung.

3. Fiskalische Tragfähigkeit

Die langfristigen, demografisch geprägten Risiken nimmt die fiskalische Tragfähigkeitsberichterstattung in den Blick. Das Bundesfinanzministerium hat mit dem Tragfähigkeitsbericht ein weit vorausschauendes Frühwarnsystem entwickelt, das sich einfügt in das vergleichbar konzipierte Berichtswesen des Wirtschaftspolitischen Ausschusses der EU sowie des Fiscal-Sustainability-Reports der EU-Kommission.⁷ Da die langfristige Tragfähigkeit der öffentlichen Finanzen die Verschuldungsmöglichkeiten determiniert, ist sie eine wesentliche Begründung für die Schuldenbremse und den europäischen Stabilitätspakt.⁸

Zentrale Funktion des finanzpolitischen Frühwarnsystems ist es, auf Gefahren für die langfristige Stabilität der öffentlichen Finanzen aus der mit dem demografischen Wandel einhergehenden Ausgabendynamik der sozialen Sicherung aufmerksam zu machen. Der Tragfähigkeitsbericht bewertet die damit verbundenen Risiken für die fiskalische

7 WIRTSCHAFTSPOLITISCHER AUSSCHUSS DER EU/EU-KOMMISSION 2018; EU-KOMMISSION 2016/ 2018. Zu den aktuellen Entwicklungen dieser Surveillance siehe die vom Bundesfinanzministerium ausgerichtete Konferenz „Fiscal Sustainability and Social Systems – Challenges and Policy Options for the Next Decade“ 2018.

8 SCHUTT, STÖßBERG 2015; BUNDESFINANZMINISTERIUM 2016 sowie allg. www.tragfaehigkeit.de.

Handlungsfähigkeit des Staates. Er fungiert somit als Stresstest, der Risikoszenarien quantifiziert und in der Konsequenz auch mögliche Politikszenarios aufzeigt.

Die Entwicklung des Frühwarnsystems bedarf der Unterstützung aus der Wissenschaft. Das ist einerseits für die Glaubwürdigkeit der Ergebnisse wichtig, andererseits der Komplexität des Systems geschuldet. Das so entwickelte Frühwarnsystem ist teilweise evidenzbasiert in dem Sinne, dass fundierte und plausible Annahmen über dem politischen Einfluss weitgehend entzogene Entwicklungen in die Modellrechnungen eingehen, diese aber ergänzt werden durch Annahmen über Größen, welche in erheblichem Maße einem politischen Einfluss unterliegen, etwa auf dem Feld der Zuwanderung. Auf Basis des Annahmensystems zeigen die langfristigen Modellrechnungen auf, welche Ergebnisse ohne korrigierendes Eingreifen der Politik zu erwarten sind („No-policy-change“-Axiom bzw. „Was-wäre-wenn“-Welt) und welche Langfristwirkungen Politikmaßnahmen im Kontext des demografischen Wandels zeitigen.

Der Tragfähigkeitsbericht kann somit aufzeigen, welcher finanzpolitische Handlungsbedarf durch den demografischen Wandel entsteht. Er zeigt also die Bürde, die bei künftigen finanzpolitischen Entscheidungen den Spielraum verengt. Da von unveränderten politischen Rahmenbedingungen ausgegangen wird und keine Anpassungsmaßnahmen angenommen werden, stellt er keine Prognose dar. Es ist Aufgabe der Finanzpolitik, durch rechtzeitige Politikentscheidungen handlungsfähig zu bleiben.

4. Spending Reviews als Bestandteil evidenzbasierter Ausgabenpolitik

International sind Spending Reviews - einnahme- und ausgabeseitige Haushaltsanalysen - in OECD-Ländern und der EU weit verbreitet. Allerdings werden sie sehr unterschiedlich interpretiert und umgesetzt. Gemeinsam ist diesen Ansätzen, dass sie der Definition von Haushaltsprioritäten, der Verbesserung der Ausgabenkontrolle, der Optimierung der Ressourcenallokation und -verwendung, der Erleichterung und Objektivierung von Neupriorisierungen oder auch der Verbesserung der Rechenschaftslegung und der Transparenz gegenüber Politik und Öffentlichkeit dienen. Im Rahmen des OECD-Netzwerks „Performance

and Results“ der Haushaltsdirektoren der OECD-Mitgliedstaaten erfolgt ein regelmäßiger internationaler Erfahrungsaustausch zu diesem Instrument, in dem für das Bundesfinanzministerium sowohl die Grundsatzabteilung als auch die Haushaltsabteilung vertreten sind.⁹

Im Zentrum der deutschen Spending Reviews steht die Verbesserung der Wirksamkeit des staatlichen Handelns. Diese selektiven Spending Reviews mit Analysen ausgewählter Politikfelder/Maßnahmen sind bisher keine „Comprehensive Reviews“ britischer Prägung, die aufgrund der dortigen Sparzwänge sehr große Aufgabenbereiche und Themenblöcke in den Mittelpunkt der Analysen stellen.

Das Instrument der Spending Reviews ist als Teil des regierungsinternen Haushaltsaufstellungsverfahrens implementiert. Über den Gegenstand der Spending Reviews entscheidet das Kabinett zum Start der Haushaltsaufstellung, Ergebnisse sollen möglichst zur Aufstellung des Haushalts des Folgejahres vorliegen.¹⁰

Abbildung 1: Beispielhafter Zeitplan eines Spending-Review-Zyklus



Die Reviews betrachten das ressortübergreifende Zusammenwirken verschiedener Politikansätze und Maßnahmen und dienen ganz konkret auch dazu, Ausgabenüberschneidungen in Folge ausgeweiteter Eigenverantwortung der Ressorts sichtbar zu machen. Die Ownership wird gestärkt, indem sich die Ressorts an der Erarbeitung und Umsetzung der Reviews beteiligen

9 Vgl. OECD 2014.

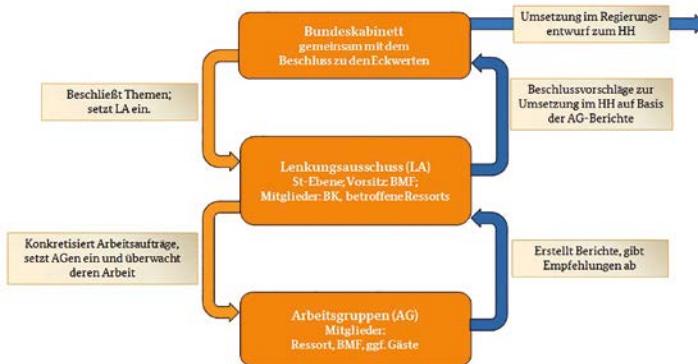
10 Vgl. zum Thema Spending Reviews BUNDESFINANZMINISTERIUM 2018a.

und die wesentlichen Ergebnisse offen gelegt werden.¹¹

Spending Reviews sind damit zunächst ein Instrument, das den Prozess der Haushaltsaufstellung um zusätzliche Informationen ergänzt. Die Nutzung dieser Informationen soll eine Neupriorisierung von öffentlichen Ausgaben bzw. die Schaffung von Spielräumen für neue Maßnahmen ermöglichen bzw. zu Verwaltungsstruktur- und Prozessverbesserungen führen, was im Ergebnis zu einer qualitativen Verbesserung der Struktur der öffentlichen Haushalte führen kann. Spending Reviews sind damit als Instrument zur Überprüfung von Zielerreichung (Effektivität) und Wirtschaftlichkeit (Effizienz) zu verstehen.

Spending Reviews sind mehr als Evaluierung. Sie dienen der Bewertung von Programmen auf Grundlage existierender Informationen und Daten, sie betrachten das Zusammenwirken mehrerer Programme und sie dienen – sozusagen als Alleinstellungsmerkmal – auch der Überprüfung der Ziele und einer ressortübergreifenden Betrachtung von Maßnahmen.

Abbildung 2: Organisationsstruktur von Spending Reviews



Folgefragen, die sich im Zuge der Reviews typischerweise stellen, lauten: Sind unterschiedliche Maßnahmen notwendig oder sollten diese konzentriert werden? Kann die Ausgabenstruktur durch eine Umschichtung verbessert werden? Erreichen einige Programme und Maßnahmen die Ziele deutlich besser als andere?

¹¹ Diese Grundprinzipien zur Erhöhung der Ownership wurde bereits früh im EU-Kontext im Rahmen der Diskussion um die Qualität der öffentlichen Finanzen betont.

Die Themen werden grundsätzlich vom Kabinett beschlossen, der Lenkungsausschuss auf der Ebene der Staatssekretäre beschließt dann den konkreten Arbeitsauftrag an eine einzusetzende Arbeitsgruppe. Die Themenvorschläge können sowohl vom Bundesfinanzministerium als auch aus den Fachressorts kommen.

In der Regel erfolgen pro Jahr zwei Spending Reviews als Teil des regierungsinternen Haushaltsaufstellungsverfahrens. Ergebnisse der Reviews sollen in der Regel im darauffolgenden Haushalt umgesetzt werden können; aufgezeigte Struktur- oder Prozessverbesserungen benötigen ggf. einen längeren Umsetzungsprozess. Mit Blick auf den Anspruch weitgehender Transparenz werden sowohl Parlament als auch Öffentlichkeit nach Abschluss über die Ergebnisse der Reviews durch Veröffentlichung im Finanzbericht und Internet informiert.¹²

Die bisher durchgeführten Spending Reviews dienten auch dazu, das Instrument zu erproben und aus Erfahrungen zu lernen. Im ersten Review-Zyklus 2015/2016 wurden zwei kleinere und eng begrenzte Förderprogramme betrachtet; nur jeweils ein Ressort (Verkehrsministerium bzw. Ministerium für Arbeit und Soziales) war beteiligt. Im zweiten Zyklus wurden mit Förderprogrammen in den Bereichen Klima und Energie sowie dem Wohnungswesen politisch und finanziell bedeutsame, ressortübergreifende Themen aufgerufen. Involviert waren in diesen Prozess bereits sechs Ressorts. Im dritten Review-Zyklus 2017/2018 wurde auf Vorschlag der beiden Ressorts die Themenblöcke „Humanitäre Hilfe“ des Auswärtigen Amtes und „Entwicklungs- und strukturbildende Übergangshilfen“ des Entwicklungshilfeministeriums sowie ein Thema der allgemeinen Verwaltung („Beschaffung standardisierter Massengüter“) einem Review unterzogen.

Nach eng abgegrenzten Maßnahmen zu Beginn liegt somit heute der Fokus auf ressortübergreifenden Reviews. Insgesamt waren 11 der 14 Ressorts der Bundesregierung an Spending Reviews beteiligt, damit ist der institutionelle Roll-out weitgehend vollzogen. Eine breitere Fundierung hat die Ressortzusammenarbeit zudem durch Fachworkshops erfahren, die den konzeptionellen Erfahrungsaustausch intensivieren.¹³

12 Siehe aktuell den Finanzbericht in BUNDESFINANZMINISTERIUM 2017b sowie BUNDESFINANZMINISTERIUM 2018a.

13 Siehe die Dokumentation BUNDESFINANZMINISTERIUM 2018b.

Die ersten Erfahrungen mit dem Instrument zeigen, dass das neue Review-System gute Ergebnisse liefert, wenn die Beteiligten ergebnisoffen mitarbeiten, die relevanten Akteure also nicht in den Schützengräben klassischer Ressortauseinandersetzungen verharren. Der Arbeitsgruppenprozess unterstützt dies durch die Bildung eines geschützten Raums, der eine ressortübergreifende Arbeit ermöglicht. Der Fokus auf Wirkungen der eingesetzten finanziellen Mittel ermöglicht eine fachliche Betrachtung ohne politische Vorgaben. Wichtig ist, im Rahmen des Auftrags genügend Gestaltungsspielraum für die Arbeitsgruppen in der konkreten Umsetzung zu lassen. Die Arbeitsgruppen beziehen Expertenwissen ein, allerdings ohne den Prozess zu sehr zu verwissenschaftlichen. Komplementär dazu garantiert der Lenkungsausschuss auf Ebene der Staatssekretäre die notwendige Entscheidungskraft.

Spending Reviews sind dann besonders effektiv, wenn sie ressortübergreifend Überschneidungen bzw. Wechselwirkungen zwischen Maßnahmen und Programmen sichtbar machen. Der analytische Blick geht dann über das einzelne Förderprogramm oder einzelne Maßnahmen hinaus. Die Einbindung in den Haushaltsaufstellungsprozess ermöglicht eine koordinierte Themenfindung und stringente Umsetzung der Ergebnisse. Sie befördert auch ein ergebnisorientiertes Arbeiten im vorgegebenen Zeitrahmen. Besonders wichtig ist in diesem Zusammenhang auch die Sicherung der Spending-Review-Ergebnisse im parlamentarischen Verfahren sowie durch fest vereinbarte gemeinsame Umsetzungskontrolle mit den Ressorts.

Gleichwohl ist in der Bewertung zu berücksichtigen, dass bisher nur ein relativ geringer Anteil der Ausgaben des Bundeshaushalts Reviews unterzogen ist. Große Budgetbereiche wurden bisher nicht bewertet. Zudem finden die Spending Reviews bisher in einem haushaltspolitisch sehr günstigen Umfeld statt, insofern stehen institutionelle Bewährungsproben noch aus.

5. Subventionsbericht als Informations- und Koordinierungsinstrument

Im Top-down-Verfahren entfällt der Verhandlungsprozess, der einen regen Informationsaustausch zwischen Fachressorts und Finanzministerium nach sich zog. Zudem kann die dezentrale Entscheidung in Eigenverantwortung der Fachressorts zu Überschneidungen der Aus-

gabenpolitiken führen. Um dieses zu vermeiden, müssen zum Beispiel Subventionspolitiken weiterhin haushalts- und finanzpolitisch koordiniert werden.

In diesem Kontext kommt dem Subventionsbericht der Bundesregierung, für den im Stabilitäts- und Wachstumsgesetz vor allem eine Rolle als Informationsinstrument für Bundestag und Bundesrat im Rahmen der Haushaltsaufstellung vorgesehen ist, eine neue Bedeutung zu.

Der Subventionsbericht als gemeinsamer Bericht der Bundesregierung, der unter Federführung des Finanzministeriums erstellt wird, übernimmt Aufgaben der Information und Koordination. Zudem berichtet er über die Umsetzung der Leitlinien der Subventionspolitik, die sich die Bundesregierung als Selbstbindung für die Subventionspolitik gesetzt hat. Er schafft Transparenz über die Aktivitäten der Ressorts sowohl für andere Fachressorts als auch für das Finanzministerium. Mit standardisierten Vorgaben für die Berichterstattung über die einzelnen Subventionen legt er eine breite Informationsgrundlage, die auch Angaben über Umfang und Ergebnisse durchgeführter Evaluierungen umfasst.

Die im Jahr 2015 eingeführte Nachhaltigkeitsprüfung für Subventionen gibt zudem der Diskussion der Auswirkungen der Subventionen auf den Kanon der Nachhaltigkeitsziele unter den Ressorts ein Forum. Auswirkungen auf Nachhaltigkeitsziele werden dokumentiert und bewertet.

Der Subventionsbericht trägt somit auch innerhalb der Bundesregierung zu einer verbesserten Informationslage bei und ermöglicht die Koordinierung teilweise naturgemäß widersprüchlicher Politikfelder.¹⁴

6. Ausblick

Spending Reviews, aber auch andere im Bundesfinanzministerium praktizierte Monitoringsysteme wie der Tragfähigkeitsbericht oder die Subventionsberichterstattung zeigen, dass ein entscheidender operativer Erfolgsfaktor in der ressortübergreifenden Durchführung und Identifikation der Verantwortungsträger mit dem Vorhaben liegt. Hierzu müssen die institutionellen Voraussetzungen geschaffen werden. Dabei ist es wichtig, die Ansätze miteinander zu verzahnen, um im Interesse

14 BUNDESFINANZMINISTERIUM 2017c.

einer effizienteren Allokation der öffentlichen Finanzen generierte Informationen bestmöglich einsetzen zu können.

Dabei gilt für eine evidenzbasierte Finanzpolitik grundsätzlich, dass der praktisch-institutionelle Rahmen und die akademisch-abstrakte Perspektive zu vereinen sind. Erst dann kann evidenzbasierte Finanzpolitik etwas bewegen.

Die idealen Instrumente der evidenzbasierten Analyse und Beratung können sich im Zeitablauf ändern, wenn sich der politische und institutionelle Rahmen ändert. Eine Frage ist z.B. wie bei einer stärker fragmentierten politischen Landschaft und gleichzeitig komplexeren Sachverhalten die Verantwortung der Politik für nachhaltige und effiziente Fiskalpolitik gestärkt werden kann.

Ein Instrument dafür wäre z.B. die Analyse und Bewertung von Wahlprogrammen durch Politikplattformen im Wahlkampf und anschließend von Koalitionsverträgen. Auf diese Weise würde für die Bürgerinnen und Bürger transparenter, was Politik kostet (und nutzt), welche Risiken eingegangen werden, und wie die Effizienz der Vorschläge/Vereinbarungen einzuschätzen ist. In den Niederlanden übernimmt diese Aufgabe das Centraal Plan Bureau CPB, das offizielle Beratungsorgan der Regierung in Wirtschafts- und Finanzfragen. In Deutschland könnte eine solche Aufgabe auch von einzelnen oder einer Gruppe von Forschungsinstituten übernommen werden. Eine Vertiefung dieses Gedankens geht aber über den Rahmen dieser Studie hinaus.

Literatur

Afonso, A., Ebert, W., Schuknecht, L., Thöne, M. (2004). „Quality of Public Finances and Growth“. ECB Working Paper 438.

Berenberg Bank (2018). „Pensions: the Next Leg of the Crisis“. Hamburg.

Bundesfinanzministerium (2016). „Vierter Bericht zur Tragfähigkeit der öffentlichen Finanzen“.

Bundesfinanzministerium (2017a). „Wirkungsorientierung im Bundeshaushalt – mehr als Spending Reviews“. BMF Monatsbericht September 2017, 1 - 6.

Bundesfinanzministerium (2017b). „Finanzbericht 2018“.

<https://www.bundesfinanzministerium.de/Content/DE/Standardarti->

kel/Themen/Oeffentliche_Finzen/Wirtschafts_und_Finanzdaten/Finanzberichte/Finanzbericht-2018-anl.pdf?__blob=publicationFile&v=1

Bundesfinanzministerium (2017c). „26. Subventionsbericht. Bericht der Bundesregierung über die Entwicklung der Finanzhilfen des Bundes und der Steuervergünstigungen für die Jahre 2015 bis 2018“.

Bundesfinanzministerium (2018a). „Spending Reviews“. https://www.bundesfinanzministerium.de/Web/DE/Themen/Oeffentliche_Finzen/Bundeshaushalt/Spending_Reviews/spending-reviews_2017.html.

Bundesfinanzministerium (2018b). „Fiscal Sustainability and Social Systems – Challenges and Policy Options for the Next Decade“. 22. Juni 2018, Dokumentation der Tagung. <http://www.bundesfinanzministerium.de/conference-documents>, Berlin.

EU-Kommission (2016). „Fiscal Sustainability Report 2015“. Institutional Paper 018. Brüssel.

EU-Kommission (2018). „Debt Sustainability Monitor 2017“. Institutional Paper 071. Brüssel.

IWF (2018). „Fiscal Monitor April 2018“. Washington D.C.

Nerlich, C., Schroth, J. (2018). „The Economic Impact of Population Ageing and Pension Reforms“. ECB Economic Bulletin, Issue 2.

OECD (2014). „OECD Principles of Budgetary Governance“. Paris.

Schuknecht, L., Zemanek, H. (2018). „Fiscal Dominance“. CESifo Working Paper Series No. 6894. München.

Schutt, R., Stoßberg, S. (2015). „Tragfähigkeit der öffentlichen Finanzen“. In: Gatzert, W., Schweisfurth, T. (Hrsg.), Öffentliche Finanzwirtschaft in der Staatspraxis, 327-356. Berlin: BWV.

Wirtschaftspolitischer Ausschuss der EU/EU-Kommission (2016). „Joint Report on Health Care and Long-Term Care Systems & Fiscal Sustainability“. Institutional Paper 37. Brüssel.

Wirtschaftspolitischer Ausschuss der EU/EU-Kommission (2018). „2018 Ageing Report of the EU“. Brüssel.

4. Politikfeld Gesundheit

Evidenzbasierung in der Medizin: Status quo und neue Herausforderungen

Von JOACHIM WINTER

Korrespondenzadresse:

Seminar für empirische Wirtschaftsforschung
Ludwig-Maximilians-Universität München
Ludwigstraße 33
80539 München
winter@lmu.de

In der Medizin ist Evidenzbasierung traditionell in mehreren Bereichen von großer Bedeutung. Am bekanntesten ist vielleicht die Zulassung von Arzneimitteln, wobei im Rahmen des Zulassungsverfahrens die Wirksamkeit, die Unbedenklichkeit und die Qualität in der Regel durch randomisierte klinische Studien belegt werden. Es handelt sich um eine klassische Anwendung von Verfahren der Evidenzbasierung. Darüber hinaus rückte in jüngerer Zeit eine Evidenzbasierung auch im Zusammenhang mit Finanzierungsfragen ins Blickfeld. Beispielsweise sind evidenzbasierte Verfahren zur Kosten-Nutzen-Bewertung von Arzneimitteln nicht Teil der Arzneimittelzulassung, sondern dienen der Beurteilung der Erstattungsfähigkeit der Kosten von Arzneimitteln, z.B. durch die Krankenkassen.

Die Verfahren der Arzneimittelzulassung und weitgehend auch diejenigen zur Bewertung der Kosten sowie die Umsetzung von Kosten-Nutzen-Abwägungen sind gesetzlich geregelt und institutionell vielschichtig verankert, so durch das Bundesinstitut für Arzneimittel und Medizinprodukte, welches u.a. für die Zulassung von Arzneimitteln zuständig ist, und das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Die methodische Ausbildung von Gesundheitsökonominnen und -ökonomern, die in diesen Bereichen tätig sind, findet auf hohem

Niveau statt. Bei den Ansprechpartnerinnen und -partnern in Politik und Verwaltung ist ein Verständnis der Bedeutung der Evidenzbasierung ausgeprägt.

In einem zweiten Sinne ist der Begriff „evidenzbasierte Medizin“ weiter zu fassen, denn alle therapeutischen Entscheidungen von Ärztinnen und Ärzten – und zunehmend auch diejenigen der Patientinnen und Patienten – sollen nach herrschendem Verständnis evidenzbasiert getroffen werden. Die Kernidee der evidenzbasierten Medizin, dass bei wichtigen Entscheidungen nicht allein auf persönliche Erfahrung und subjektive Einschätzungen zurückgegriffen, sondern der Stand der aktuell besten verfügbaren wissenschaftlichen Evidenz berücksichtigt werden sollte, wird breit unterstützt.¹ Sie findet ihren Niederschlag in Initiativen zur systematischen Sammlung der Daten klinischer Studien, ihrer Aufarbeitung in Form von Metastudien und in der Formulierung von Behandlungsleitlinien.² Hier gibt es jedoch kaum gesetzliche Vorgaben. Auch Finanzierungsfragen werden oft ausgeblendet – im Vordergrund steht bei der Entwicklung von Metastudien und Leitlinien die medizinische Wirksamkeit von Medikamenten und Therapien.

Der dritte Bereich, in dem Evidenzbasierung zunehmend wichtig wird, ist im Kern ein ökonomischer: Für die Gesundheitspolitik von zentraler Bedeutung ist die Wirkung der institutionellen Anreize (insbesondere derjenigen, die von den Finanzierungssystemen ausgehen) auf die Behandlungsentscheidungen der Ärztinnen und Ärzte und anderer Leistungserbringer. Ebenso wichtig ist das Verständnis der gesundheitsbezogenen Verhaltensweisen von gesunden Menschen und Patientinnen und Patienten, beispielsweise bei der Prävention und der Therapieadhärenz, aber auch in anderen Bereichen wie der Wahl von Krankenversicherungsverträgen.

An einigen Stellen hat der Gesetzgeber auch in diesen weiteren Bereichen die Idee der Evidenzbasierung zumindest in Ansätzen umgesetzt. Als Beispiel sei die Einführung diagnosebezogener Fallgruppen (Diagnosis Related Groups, DRGs) bei der Vergütung medizinischer Leistungen in Krankenhäusern mit dem Gesundheitsreformgesetz vom 22.12.1999

1 Diese Sichtweise wurde insbesondere in einem Beitrag von SACKETT et al. 1996 etabliert; das Konzept einer evidenzbasierten Medizin hat allerdings eine lange Geschichte.

2 Exemplarisch sei hier die internationale Cochrane Collaboration genannt; siehe für Deutschland <https://www.cochrane.de/de>.

genannt. In §17b des Krankenhausfinanzierungsgesetzes (KHG) wurde die Selbstverwaltung aufgefordert, ein international erprobtes DRG-System in Deutschland einzuführen. Am 27.06.2000 hat sich die Selbstverwaltung für das australische AR-DRG-System als Grundlage für ein zu entwickelndes deutsches System entschieden. Das deutsche G-DRG-System wurde unter Berücksichtigung bundesdeutscher Leistungs- und Kostendaten entwickelt. Jährliche Neukalkulationen sollen die Anpassung des Systems an die Behandlungswirklichkeit gewährleisten. Die Kodierung von Diagnosen und Leistungen ist Grundvoraussetzung für die Zuordnung des einzelnen Krankenhauses zu einer DRG-Fallpauschale und selbstverständlich muss die Kodierung einheitlichen Regeln folgen. Vor diesem Hintergrund fordert der Gesetzgeber eine regelmäßige Weiterentwicklung des DRG-Systems. Die entsprechenden evidenzbasierten Verfahren und die ausführenden Institutionen sind inzwischen etabliert, zum Beispiel durch das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI).

Nicht gesetzlich geregelt ist allerdings, wie überprüft werden soll, ob von der Gestaltung eines DRG-Systems Fehlanreize ausgehen, die die Leistungserbringer zu Manipulationen durch sachlich nicht zu rechtfertigende Hochkodierung („Upcoding“) verleiten. Dafür gibt es Evidenz. Ein Beispiel ist Hochkodierung bei der Vergütung von Krankenhäusern für Frühgeburten. Die Fallpauschalen hängen vom Geburtsgewicht ab und steigen beim Unterschreiten bestimmter Schwellenwerte (2500 Gramm, 2000 Gramm etc.) jeweils deutlich an. Dies führt dazu, dass wesentlich häufiger Geburtsgewichte angegeben werden, die knapp unterhalb dieser Schwellen liegen, als statistisch zu erwarten wäre. Dies führt zu höheren Vergütungen im Rahmen des DRG-Systems.³

Festzuhalten bleibt: Evidenzbasierung besitzt traditionell große Bedeutung und auch hohe Akzeptanz in der Medizin – zumindest in einem engeren Sinne. Sie ist vielfältig gesetzlich und institutionell verankert. Dennoch gibt es Defizite, vor allem dort, wo grundlegende ökonomische Mechanismen, Verhaltensweisen von Patientinnen und Patienten, Ärztinnen und Ärzten und anderen Leistungserbringern sowie die Wechsel-

3 Upcoding von Geburtsgewichten für Frühgeburten wurde anhand deutscher Daten von JÜRGES, KÖBERLEIN 2015 sowie von REIF, WICHERT, WUPPERMANN 2018 nachgewiesen. Auch international gibt es umfangreiche Evidenz zum Upcoding bei diesen und anderen DRGs.

wirkung mit den sozioökonomischen Lebensumständen von Bedeutung sind. Die Einsicht, dass Evidenzbasierung auch in diesen Bereichen notwendig und möglich ist, scheint vielen Akteuren noch fremd.

Die randomisierte Kontrollstudie ist als methodischer „Goldstandard“ fest in der medizinischen Fachcommunity verankert. Andere Verfahren der Evidenzgewinnung – zum Beispiel aufgrund von Daten aus sozialwissenschaftlichen Beobachtungsstudien oder Abrechnungsdaten der Krankenkassen und durch Ausnutzung sogenannter natürlicher Experimente, wenn echte randomisierte Studien nicht durchführbar sind – werden vielfach nicht akzeptiert. Die Ergebnisse der genannten Studien zu Fehlanreizen im DRG-System wären durch eine randomisierte Kontrollstudie aber nicht erzielbar gewesen; stattdessen wurden ökonometrische Verfahren eingesetzt, die sogenannte „natürliche“ Experimente ausnutzen. Während solche Verfahren in der Gesundheitsökonomik etabliert sind und auch in Deutschland auf hohem Niveau in wissenschaftlichen Studien eingesetzt werden, sind sie in Bereichen, in denen die randomisierte Kontrollstudie als alleiniger Goldstandard gilt, nicht akzeptiert.

Ganz neue Möglichkeiten zur Gewinnung kausaler Evidenz im Gesundheitsbereich ergeben sich schließlich aus der Nutzung von Massendaten („Big Data“), insbesondere von Abrechnungs- und Diagnosedaten, in Kombination mit statistischen Verfahren des maschinellen Lernens („Machine Learning“). Eine Reihe neuer Arbeiten zeigt, dass solche Ansätze nicht nur traditionelle Methoden ergänzen, sondern ganz neue Fragestellungen erschließen können. Hier bedarf es der wissenschaftlichen Neugier, aber auch der methodischen Weiterentwicklung der Gesundheitsökonominnen und -ökonom.

Defizite bestehen auch beim Datenzugang, der Voraussetzung für die Anwendung moderner Methoden in der evidenzgenerierenden Forschung ist. Der Zugang zu Abrechnungs- und Kostendaten für Zwecke der gesundheitsökonomischen Forschung ist nach wie vor umständlich und aufwändig. Die statistischen Ämter sind oft, aber nicht immer hilfsbereit. Untersuchungen zu den langfristigen Wirkungen von Therapien und Medikamenten, aber auch von Krankheitsverläufen können an den Löschungsaufgaben nach §75 SGB X scheitern. Die Verknüpfung von Prozessdaten (z.B. Abrechnungsdaten) mit Daten aus sozialwissenschaftlichen Befragungsstudien wie dem Sozio-ökonomischen Panel (SOEP) und dem Survey of Health, Aging and Retirement in Europe (SHARE) steht

vor nahezu unüberwindlichen rechtlichen Hindernissen. Bei der Rentenversicherung in Deutschland und bei der Krankenversicherung in den USA (wo regelmäßig eine Verknüpfung von Medicare-Abrechnungsdaten mit Befragungsdaten der Health and Retirement Study, HRS, erfolgt) ist eine solche Verknüpfung möglich. Verschiedene Initiativen, unter anderem des Rats für Sozial- und Wirtschaftsdaten (RatSWD) bemühen sich, die unbefriedigende Situation beim Datenzugang zu verbessern, damit die aktuellen Möglichkeiten einer evidenzbasierten Medizin und Gesundheitsökonomik auch in Deutschland voll umgesetzt werden können.

Literatur

Hernán, M.A., Robins, J.M. (2016). „Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available”. *American Journal of Epidemiology* 183(8), 758–764.

Jürges, H., Köberlein, J., (2015). „What Explains DRG Upcoding in Neonatology? The Roles of Financial Incentives and Infant Health”. *Journal of Health Economics* 43, 13–26.

Khoury, M.J., Ioannidis, J.P.A. (2014). „Big Data Meets Public Health”. *Science* 346, 1054–1055.

Mooney, S.J., Westreich, D.J., El-Sayed, A.M. (2015). „Commentary: Epidemiology in the Era of Big Data.” *Epidemiology* 26(3), 390–394.

Obermeyer, Z., Emanuel, E.J. (2016). „Predicting the Future — Big Data, Machine Learning, and Clinical Medicine”. *New England Journal of Medicine* 375, 1216–1219.

Reif, S., Wichert, S., Wuppermann, A. (2018). „Is it good to be too light? Birth weight thresholds in hospital reimbursement systems”. *Journal of Health Economics* 59, 1-25.

Sackett, D.L., Rosenberg, W.M.C., Muir Gray, J.A., Haynes, R.B., Richardson, W.S. (1996). „Evidence Based Medicine: What it is and What it isn't”. *BMJ* 312, 71-72.

Schneeweiss, S. (2014). „Learning from Big Health Care Data”. *New England Journal of Medicine* 370, 2161– 2163.

Shiffrin, R.M. (2016). „Drawing Causal Inference from Big Data”. *Proceedings of the National Academy of Sciences of the United States of America* 113, 7308–7309.

Evidenzbasierung in Medizin, Gesundheitspolitik und -versorgung: Drei Thesen

Von REINER LEIDL

Korrespondenzadresse:

Ludwig-Maximilians-Universität München

Institut für Gesundheitsökonomie und Management im Gesundheitswesen

Ludwigstr. 28

80539 München

leidl@bwl.lmu.de

Ab Beginn der 1990er Jahre hat im Gesundheitssystem evidenzbasiertes Handeln unter dem Stichwort ‚Evidence-based Medicine‘ einen ungeheuren Aufschwung erfahren.¹ Medizinische Entscheidungen sollten auf einem wissenschaftlichen Nachweis der zu erwartenden Effektivität beruhen. Neben klinischen Maßnahmen wurde das Konzept auch für Maßnahmen im Public-Health-Bereich, etwa bei der Auswahl von Strategien zur Prävention des Rauchens, eingesetzt. Medizinisches Handeln, dazu gehörige Organisationen und Hilfsmittel, aber auch die Forschung und die Wissensbasis wurden durch dieses Konzept substantiell beeinflusst und weiterentwickelt.² Dazu will ich im Folgenden drei Thesen darlegen.

1. These: Produkt- bzw. Technologieorientierung hat die evidenzbasierte Medizin zum Erfolg geführt.

Medizinisches Handeln bedeutet, Entscheidungen zu treffen über Maßnahmen zur Versorgung von Patientinnen und Patienten, etwa zur Diagnostik, Prävention, Therapie oder Rehabilitation. Solche Maßnahmen lassen sich als Technologien verstehen, die bei der Versorgung zum Einsatz kommen. Beim Einsatz von Arzneimitteln kann man auch von Produkten sprechen. Zum Wohl der Patientinnen und Patienten verlangt die evidenzbasierte Medizin für die Entscheidung zum Technologieeinsatz einen wissenschaftlichen Effektivitätsbeleg.

1 Evidence-Based Working Group 1992.

2 DJULBEGOVIC 2017.

In der Medizin rief dieses Konzept durchaus heftige Kritik hervor, etwa an einer zu starken Vereinfachung, Anwendung von „Kochbuchmedizin“ und Ignorierung der ärztlichen Intuition.³ Dennoch haben zahlreiche Faktoren die weitgehende Umsetzung dieses Konzepts begünstigt. So sind in der Medizin in den großen Indikationsgebieten viele ähnliche Entscheidungen zu ähnlichen Krankheitsproblemen zu treffen. Die Anforderung, dass Maßnahmen effektiv sein müssen, ist in einigen sozialstaatlichen Systemen sogar gesetzlich kodifiziert, etwa im § 12 des Sozialgesetzbuchs V zur Wirtschaftlichkeit in Deutschland. Die Wirksamkeit der einzusetzenden Produkte und Technologien lässt sich häufig gut experimentell prüfen, womit der naturwissenschaftliche Nachweisstandard anwendbar wird. Der wissenschaftliche Nachweis bedient dann auch die Erwartung von Beteiligten wie der Krankenkassen, die Transparenz über die Wirksamkeit neuer Technologien fordern.

Ihre eigenen Ansprüche und die externen Anforderungen konnte die Medizin als ein sehr großes und leistungsfähiges akademisches Fach schon bald immer besser erfüllen. So verzeichnet die PubMed-Datenbank, eine zentrale und öffentlich frei zugängliche Wissensquelle der Lebenswissenschaften, Ende Mai 2018 insgesamt knapp 800.000 Arbeiten, die der Kategorie ‚klinische Versuche‘ zugeordnet wurden. Das ist mehr als das 42-fache der Zahl an Studien, welche in der Ökonomie von der – nicht-öffentlichen – Econlit-Datenbank zu Stichworten zu experimenteller Forschung wiedergegeben wird. Dazu kommt eine standardisierte Wissensdokumentation: In der PubMed-Datenbank sind viele Arbeiten in strukturierten Abstracts zusammengefasst, mit denen sich Ziel, Vorgehensweise und Resultat eines Beitrags oft treffend einordnen lassen. Auch gesundheitsökonomische Evaluationen werden in dieser Datenbank erfasst.

Eine wesentliche Stütze der Evidenz in der Medizin sind die systematischen Literaturübersichten, für die anerkannte Qualitätsstandards entwickelt wurden.⁴ Diese Arbeiten verdichten mit transparenten Suchen und Auswertungen das verfügbare Wissen aus vielen vergleichbaren Studien und synthetisieren damit die Evidenz. Ende Mai 2018 verzeichnet die PubMed-Datenbank mehr als neunmal so viele Arbeiten

3 DJULBEGOVIC 2017.

4 MOHER 2009.

mit ‚Review‘ im Titel als die Econlit-Datenbank. Dabei lag der Anteil der Arbeiten mit ‚Systematic Review‘ im Titel bei gut über einem Viertel, während er in der ökonomischen Datenbank nicht einmal ein Prozent erreichte.

Die insbesondere durch klinische Versuche generierte Evidenz wirkt prägend für die Qualitätsstandards des medizinischen Handelns selbst, die in Leitlinien niedergelegt werden. Soweit möglich, sollen diese auf Evidenz und Expertenkonsens basieren. In Deutschland koordiniert die Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) die Arbeit ihrer 178 Fachgesellschaften an medizinischen Leitlinien. 2017 wurden 757 Leitlinien in drei evidenzbezogenen Qualitätsklassen auf den Webseiten dieser Arbeitsgemeinschaft gelistet. Leitlinien als Transmissionsmechanismus wissenschaftlicher Ergebnisse in konkrete Handlungsempfehlungen stellen damit eine zentrale Stütze der evidenzbasierten Medizin dar.⁵ Mit der strukturierten, praxisrelevanten Anwendung steigen zugleich die Anforderungen an Umfang und Qualität der Evidenzbasis.

Ein weiterer Transmissionsmechanismus setzt die Evidenz zur Steuerung des Einsatzes neuer Technologien ein. Strenge, primär auf randomisierten klinischen Versuchen basierende Zulassungsverfahren und weitere Anforderungen an den Nachweis von Patientennutzen und Wirtschaftlichkeit des Technologieeinsatzes sind für den Marktzugang in Systemen der sozialen Sicherung zu erfüllen. Dies hat sich über Jahrzehnte im Bereich der Arzneimittelversorgung etabliert und wird in jüngster Zeit auch auf den Bereich der Medizintechnik und Medizinprodukte ausgedehnt. Diese regulatorischen Anforderungen wirken auf die Qualität und Vergleichbarkeit der dafür durchgeführten klinischen Versuche ein. Das gilt auch für daran anknüpfende ökonomische Bewertungen, die in manchen Ländern für Erstattungsentscheidungen herangezogen werden.⁶ Dabei kommen die verschiedenen Jurisdiktionen oft zu deutlich unterschiedlichen Entscheidungen.⁷

Diese institutionellen Einbindungen etablieren Qualitätsanforderungen und eine feste Schnittstelle zur Umsetzung der Evidenz in

5 DJULBEGOVIC 2017.

6 PANTELI ET AL. 2015.

7 FISCHER 2016.

Entscheidungen. Da eine öffentlich dokumentierte Anhörung von Beteiligten in vielen Regulierungsprozessen als politisch erforderlich angesehen wird, gibt es bei der Bewertung der Evidenz meist eine hohe Transparenz. Eine Ausnahme davon bildet in einer Reihe von Jurisdiktionen die Festlegung der Erstattungshöhe für eine Technologie. Auch wenn Beteiligte und Betroffene teils deutliche Kritik an den Regulierungsentscheidungen üben, können diese eine bemerkenswerte Stabilität erreichen. So wurde das englische National Institute for Health and Care Excellence 1999 gegründet und hat mit gewissen Modifikationen seither eine ganze Reihe politisch unterschiedlicher Administrationen überstanden. Die aktuelle deutsche Regulierung des Zugangs für Arzneimittel zum Markt der gesetzlich Versicherten stammt aus 2011 und erscheint ebenfalls fest etabliert.

2. These: Gesundheitspolitik auf Evidenz zu basieren ist ebenso schwer wie bei anderen Politiken

Gesundheitspolitik kann national, regional oder sektoral abgegrenzt sein, aber meist zeichnet sie sich wie andere Politikfelder durch hohe Komplexität aus. Dazu können spezifische Faktoren beitragen wie die Verflechtung von Gesundheitsversorgung und Krankenversicherung, aber auch generelle Faktoren wie konfligierende Interessenlagen der Beteiligten, die mit zahlreichen Informationsasymmetrien und Unsicherheiten über mögliche politische Maßnahmen und deren zu erwartende Effektivität verbunden sein können. Unklare Zielsetzungen und eine hohe Regulierungsdichte und -frequenz können zusätzlich die Beurteilung einzelner politischer Maßnahmen erschweren.⁸

Im Vergleich zur Technologiebewertung aus These 1 lässt sich eine Evaluation der Gesundheitspolitik wie folgt charakterisieren: Die politischen Maßnahmen erfolgen oft in historisch einmaligen Situationen, und sie haben ein meist breiteres Zielspektrum und vielfältigere Erfolgskriterien als Technologien. Für ihre Beurteilung können kaum experimentelle Situationen herangezogen werden, vielmehr dominieren Beobachtungsstudien mit den ihnen eigenen Einschränkungen, insbesondere der Gefahr von Verzerrungen in der Ursachenanalyse.

Die Evaluation kann Teil der Politik sein und sich damit außerhalb

8 Vgl. etwa zu Kostendämpfungsstrategien LEIDL 1997.

einer unabhängigen wissenschaftlichen Beurteilung positionieren. Zu den Politikwirkungen können zudem Marktreaktionen, öffentliche Meinungsäußerungen und Wählerreaktionen gehören, was die Wirkungsdimensionen weiter auffächert. Im Vergleich zur Technologiebewertung erscheint die Wissensbasis zur Politikbewertung deutlich heterogener, weniger strukturiert und bezüglich einer Qualitätsbeurteilung weniger leicht zugänglich. Dies alles erschwert auch eine systematische Aufbereitung der Evidenz. Schließlich kann die Transparenz von Politikprozess und -wirkungen zu Zielkonflikten führen. So verwundert es wenig, dass zur wissenschaftlichen Evidenzbasierung auch der Gesundheitspolitik nur vergleichsweise wenige Beispiele zu finden sind. Gesundheitspolitik erscheint somit keineswegs leichter evaluierbar als andere Politikfelder und eine Evidenzbasierung der Politik dürfte somit auf grundsätzlich vergleichbare Schwierigkeiten wie in anderen Politikbereichen treffen. Insgesamt ist zu vermuten, dass evidenzbasierte Gesundheitspolitik erst um Akzeptanz ringen muss und dieses als eine besonders anspruchsvolle Aufgabe zu werten ist.

3. These: Komplexe Managementansätze können in der Gesundheitsversorgung auf Evidenz gestützt werden

Neben den beiden Kontrapunkten in der Evidenz im Gesundheitsbereich, den experimentell prüfbar Technologien und den schwer evaluierbaren komplexen Politiken gibt es noch eine intermediäre Evaluationsebene: Umfassende Ansätze zum Management von chronischen Krankheiten verbinden Elemente aus beiden vorgenannten Bereichen, nämlich über einen langen Zeitraum zahlreiche Maßnahmen, darunter viele ‚Technologien‘, mit der übergreifenden Zielsetzung, die Lebensdauer und insbesondere die Lebensqualität zu verbessern. Diese umfassenden Ansätze werden als ‚Disease Management Programme (DMP)‘ bezeichnet. Ihr problemorientierter Ansatz kann Evidenz aus dem Bereich des Technologiemanagements heranziehen; er kann aber auch selbst einer Evidenzprüfung unterzogen werden.⁹

Mit den Zielen einer Qualitätssteigerung und Kostenkontrolle sind in der deutschen gesetzlichen Krankenversicherung mehrere Millionen Versicherte in DMP eingeschrieben, etwa im Bereich Diabetes,

9 LEIDL 2013.

Herz- oder Lungenerkrankungen. Krankenkassen und Leistungsanbieter schließen dazu zahlreiche Versorgungsverträge ab. Die vielfältigen Maßnahmen betreffen sowohl verschiedene Komponenten der ärztlichen Versorgung als auch des Selbstmanagements der Patientinnen und Patienten. Die von beiden Seiten zu beachtenden Versorgungskriterien sind in der Risikostrukturausgleichsverordnung festgelegt. Zahlreiche Beteiligte, Prozesselemente, Versorgungskriterien und Ergebniskomponenten tragen zu einem hohen Komplexitätsgrad bei. Allerdings werden diese Managementansätze bei sehr vielen ähnlichen Anwendungsfällen – den Versicherten – eingesetzt und enthalten oft Einzelkomponenten, für die Evidenz aus der Technologiebewertung vorliegt. Auch die Zielvorgaben und Erfolgskriterien ähneln denjenigen aus der Technologiebewertung. Neben den Gesamteffekten können auch Wirkungen der Einzelkomponenten in der realen Versorgung und die Prozessabläufe untersucht werden. Die Wirkungen der Einzelkomponenten können teils mit experimentellen Studienergebnissen und systematisch aufbereiteter Evidenz aus der Technologiebewertung verglichen werden. Da sich das Krankheitsmanagement auf die gesamte reale Versorgung bezieht, wird es methodisch meist in Beobachtungsstudien evaluiert. In einigen Fällen werden zur Erfassung der langfristigen Wirkung von multiplen Interventionen mathematische Modelle eingesetzt, etwa entscheidungsanalytisch aufbereitete Markov-Modelle. Die Beobachtungsdaten stammen oft aus den umfänglichen, langfristigen Leistungs- und Abrechnungsdokumentationen der Krankenkassen, teils ergänzt durch die medizinischen Daten der DMP-Dokumentation und zusätzliche Fragebogenerhebungen. Zur Vermeidung von Verzerrungen können statistische Verfahren zur Kontrollgruppengenerierung eingesetzt werden, allerdings mit dem möglichen Nachteil, dass nur Ausschnitte des realen Versorgungsgeschehens beurteilt werden.

Obwohl Evaluationen von DMP gesetzlich generell vorgesehen sind, mangelt es häufig an geeigneten Studiendesigns, um dies in wissenschaftlich akzeptabler Qualität zu erreichen. Dennoch finden sich in dieser intermediären Evaluationsebene mittlerweile eine ganze Reihe wissenschaftlich hochwertiger Studien. Deren Resultate wurden auch in systematischen Übersichten aufbereitet, etwa zum Management von

Diabetes¹⁰ oder zur gesundheitsökonomischen Evaluation von Multi-komponentenansätzen in DMP.¹¹ Allerdings gilt der Wirksamkeit von DMP keine vergleichbare öffentliche Aufmerksamkeit wie dem Marktzugang etwa von neuen Arzneimitteln. Dies mag mit der Anforderung zusammenhängen, dass DMP langfristig Verbesserungen des Gesundheitsverhaltens anstreben, ein in Umsetzung und Evaluation mühevoller Bereich. Dennoch liegt zur Beurteilung von komplexen Managementansätzen relevante Evidenz vor, die für die Verbesserung von DMP genutzt werden kann. Dies bildet zugleich einen möglichen Transmissionsmechanismus in die Praxis ab. Ansätze zur Weiterentwicklung des Managements chronischer Krankheiten auf Basis wissenschaftlicher Evidenz können somit auf Akzeptanz hoffen.

Literatur

Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) e.V. <http://www.awmf.org/awmf-online-das-portal-der-wissenschaftlichen-medizin/awmf-aktuell.html> [Zugriff 04.06.2018]

Djulgovic, B., Guyatt, G.H. (2017). „Progress in Evidence-Based Medicine: A Quarter Century on”. *Lancet* 390 (10092), 415-423.

Evidence-Based Medicine Working Group (1992). „Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine”. *Journal of the American Medical Association* 268:2420e5.

EconLit Datenbank <https://www.ebsco.com/products/research-databases/econlit> [Zugriff 27.05.2018]

Fischer, K.E., Heisser, T., Stargardt, T. (2016). „Health Benefit Assessment of Pharmaceuticals: An International Comparison of Decisions from Germany, England, Scotland and Australia”. *Health Policy* 120:10, 1115-1122.

Fuchs, S., Henschke, C., Blumel, M., Busse, R. (2014). „Disease Management Programs for Type 2 Diabetes in Germany: a Systematic Literature Review Evaluating Effectiveness”. *Deutsches Ärzteblatt International* 111:26, 453-463.

10 FUCHS ET AL. 2014.

11 KIRSCH 2016.

Kirsch, F. (2016). „Economic Evaluations of Multicomponent Disease Management Programs with Markov Models: A Systematic Review”. *Value in Health* 19:8, 1039-1054.

Panteli, D., Eckhardt, H., Nolting, A., Busse, R., Kulig, M. (2015). „From Market Access to Patient Access: Overview of Evidence-Based Approaches for the Reimbursement and Pricing of Pharmaceuticals in 36 European Countries”. *Health Research, Policy and Systems* 13:39.

PubMed Datenbank <https://www.ncbi.nlm.nih.gov/pubmed>
[Zugriff 27.05.2018]

Leidl, R. (2013). „Managing Technologies or Diseases? Two Strategic Approaches to Health Care in the Ageing Population”. In: Cheung, F.M., Woo, J., Law, C. (Hrsg.), *Health Systems: Challenges, Visions, and Reforms from a Comparative Global Perspective*. 21st Century Book Series. Hong Kong Institute of Asia-Pacific Studies, The Chinese University of Hong Kong & Chinese University Press, 83-107.

Leidl, R. (1997). „Kostendämpfungspolitik in den Niederlanden und in Deutschland: eine Untersuchung von Maßnahmen in drei Leistungsbereichen”. *Zeitschrift für Gesundheitswissenschaften* 5:2, 101-122.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.; PRISMA Group (2009). „Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement”. *British Medical Journal* 339:b2535.

Evaluierung von Finanzmarktreformen: Lehren aus anderen Politikbereichen - Beispiel Gesundheit

Von J.-MATTHIAS GRAF VON DER SCHULENBURG

Korrespondenzadresse:

Center for Health Economics Research (CHERH)

Leibniz Universität Hannover

Otto-Brenner-Str. 1

30159 Hannover

jms@cherh.de

Zusammenfassung

Seit Jahrzehnten sind ökonomische Evaluationsstudien, hier mit dem Sammelbegriff „Kosten-Nutzen-Analysen (KNA)“ bezeichnet, ein Instrument zur Entscheidungsunterstützung beim Einsatz öffentlicher Mittel. Sie sind bei allen großen Investitionsvorhaben vorgeschrieben. Auch im Gesundheitswesen spielt die KNA in ihren vielfältigen Ausprägungen international eine Rolle. Sie ist auch für den Bereich der Gesetzlichen Krankenversicherung in Deutschland gesetzlich verankert (§§ 35a und b SGB V). Während es einen breiten Konsens in der Messung der Kosten gibt, sind die Verfahren zur Messung des Nutzens in Deutschland umstritten. Einen relativen Vergleich der Kosten und Nutzen wagt man deshalb selten. Dies ist ein Grund, dass von der KNA im deutschen Gesundheitswesen weniger Gebrauch gemacht wird, als dies in der Regel im Ausland üblich ist. Damit verzichtet man in Deutschland auf ein wichtiges Instrument zur Vorbereitung von Allokationsentscheidungen und der ökonomischen Bewertung medizinischer Maßnahmen und Programme (z.B. Impfprogramme oder Disease-Management-Programme). Die Methoden der KNA basieren auf Werturteilen u.a. der allgemeinen Wohlfahrtsökonomie. Man sollte sich stets bei Bewertungen der zugrundeliegenden Werturteile bewusst sein und sie offenlegen. Die Übertragung der Methoden der KNA im Gesundheitswesen auf andere Politikbereiche ist möglich und auch sinnvoll, da gerade die Gesundheitsökonomie seit Jahrzehnten auf diesem Gebiet forscht. Die Anwendung der KNA in anderen Politikbereichen mag auch einfacher sein, da es nicht um die Bewertung von menschlichem Leid geht. Oder doch?

1. Problemstellung

Ökonomische Evaluationsstudien, hier kurz mit Kosten-Nutzen-Analysen (KNA) bezeichnet, sind seit Jahrzehnten ein Instrument zur Entscheidungsfindung bei dem Einsatz öffentlicher, das heißt kollektiv aufgebracht Mittel. So schreibt beispielsweise § 7 Abs. 2 der Bundeshaushaltsordnung (BHO) Wirtschaftlichkeitsuntersuchungen für alle größeren Projekte vor, wobei die KNA explizit genannt werden (§ 7 Abs. 3 BHO). Analoge Bestimmungen finden sich in den Landeshaushaltsordnungen und in den Bestimmungen supranationaler Organisationen. Wenn ein großes Entwicklungshilfeprojekt durchgeführt werden soll, werden zur Entscheidungsunterstützung KNA durchgeführt. Denn für die knappen Mittel gibt es Ausgabenalternativen.

Andersherum, wenn Entscheidungen „alternativlos“ sind, wie es manchmal von den Entscheidungsträgern suggeriert wird, benötigt man keine Evaluation und damit keine KNA. Leistungsfähige Methoden für die Ermittlung von Kosten und Nutzen sind in 60-jähriger Forschung entwickelt worden und gehören seit langem zum finanzwissenschaftlichen Lehrbuchwissen.¹

Selbst für heikle Themen,

- wie der Wert des Lebens (zum Beispiel, wenn durch den Ausbau eines Verkehrsweges die Sicherheit erhöht und damit die Anzahl der Verkehrstoten vermindert wird oder ein Impfprogramm die Zahl der Infektionstoten senkt),
- wie in der Zukunft liegende Kosten und Nutzen (Diskontierung), die Behandlung von Risiko und Unsicherheit (Monte-Carlo-Simulationen, stochastische Modellierung),
- wie die Berücksichtigung von Kriterien der sozialen Wünschbarkeit und Verteilungsgerechtigkeit (Verteilungsgewichte) und
- wie die Erfassung schwer messbarer (intangibler) Kosten und Nutzen hat die Forschung Methoden entwickelt und bietet sie an.²

Eine besondere Schwierigkeit besteht, wenn verschiedene Kosten- und Nutzen-Dimensionen bei verschiedenen Personengruppen anfallen. Zum Beispiel bringt der Bau eines neuen Flugplatzes Zeitersparnisse für Reisende, aber auch Behinderungen während des Baus, Luftverschmut-

1 DASGUPTA, PEARCE 1972; LITTLE, MIRRELESS 1969.

2 SCHÖFFSKI, SCHULENBURG 2011.

zung und Lärm für Anlieger sowie Baukosten für die Steuerzahlenden. Das Gleiche gilt im Gesundheitswesen. Eine neue Operationsmethode mag Leben retten, aber auch zu erhöhten Operationsrisiken führen. Die Verteilung der knappen Organe in der Organtransplantation ist gut - für den, der ein Organ erhält. „The decision-maker is forced to weigh the gains and losses by his own preferences. The real danger to avoid is ignoring gains and losses simply because they cannot be valued.”³

Aufbauend auf diesen Erkenntnissen hat der Einsatz von KNA im Gesundheitswesen Einzug gehalten und ist mittlerweile im Ausland, insbesondere in Ländern mit einem steuerfinanzierten Gesundheitssystem, etabliert. Aber auch in Deutschland mit seinem gegliederten Krankenversicherungssystem hat sich die KNA als Methode zur Entscheidungsfindung im Sozialgesetzbuch V (§ 35b SGB V) festgehalten und durch die frühe Nutzenbewertung von Arzneimittelinnovationen (§35a SGB V) etabliert.

Während die Methoden der Kostenermittlung und -bewertung weitgehend akzeptiert sind (man kann ja alles schön in Geldeinheiten ausdrücken), werden die Methoden der Ermittlung und Bewertung der Nutzen von Gesundheitsleistungen kontrovers beurteilt. Im Folgenden werden wir auf die Kosten- und die Nutzenbewertung eingehen und dann einige der Werturteile benennen, die der KNA im Gesundheitswesen zugrunde liegen.

Ein Schwerpunkt dieses Beitrages liegt in der Herausarbeitung von Paradigmen, Werturteilen und Annahmen, auf denen die Methoden der KNA im Bereich des Gesundheitswesens basieren und die den Entscheidungsträgern bewusst sein sollten, wenn sie die Ergebnisse solcher Studien bei der Entscheidungsfindung berücksichtigen. Nicht eingegangen wird auf einzelne Formen und methodische Aspekte der KNA, da diese anderenorts breit dargestellt sind.⁴ Auch werden nicht die Lehren des Einsatzes der KNA im Gesundheitswesen für die Evaluation von Finanzmarktreformen diskutiert. Sicherlich ist die KNA bei der Evaluation von Finanzmarktreformen möglich. Aber die Methoden, Instrumente und die Methoden der Datenerfassung und -auswertung bedürfen einer Anpassung.

3 DASGUPTA, PEARCE 1972.

4 SCHÖFFSKI, SCHULENBURG 2011.

2. Bewertung der Kosten

Ein Teil der KNA ist die Erfassung der Kosten der zu evaluierenden Gesundheitsleistung oder des Gesundheitsprogramms sowie der möglichen Alternativen. Zu beachten ist dabei, dass es nicht Aufgabe der Ökonomik ist, Entscheidungen über die Allokation der öffentlichen Mittel zu treffen, sondern dies ist Aufgabe der dazu legitimierten Entscheidungstragenden. Die Ökonomik liefert mit ihren Methoden und Analysen nur Unterstützung. Deshalb obliegt es auch den Entscheidungstragenden zu bestimmen, aus welcher Perspektive die Kosten erfasst werden, welche Kostenarten berücksichtigt werden und wie die Kosten zu bewerten sind.

Ohne die Festlegung der Perspektive ist eine Erfassung und Bewertung der Kosten und Nutzen nicht möglich. Es kann sogar sein, dass die Kosten für den einen Nutzen für den anderen darstellen. In der Krankenversicherung stellen die Beiträge Kosten für den Versicherten dar. Sie werden verwendet für Gesundheitsleistungen der Patientinnen und Patienten, für die sie Nutzen bedeuten. Volkswirtschaftlich gesehen sind die Kosten einer Maßnahme häufig der Nutzen einer alternativen Maßnahme, die stattdessen nicht ergriffen werden konnte (Alternativkostenkonzept). Als Perspektive wird der Blickwinkel bezeichnet, aus dem die Kosten erfasst und gemessen werden bzw. wessen Kosten berücksichtigt werden. Die gesellschaftliche Perspektive versucht, die volkswirtschaftlichen Ressourcenverbräuche zu ermitteln. Reine Spillover-Effekte (Verteilungseffekte), die sich ergeben, saldieren sich in der gesellschaftlichen Perspektive zu Null. Die Sozialversicherungsperspektive, die im deutschen GKV-System vorgeschrieben ist, erfasst bei den Kosten nur die Ausgaben der Sozialversicherung. Da die Dienstleister im Gesundheitswesen nicht nach Marktpreisen bezahlt werden, sondern ihre Leistungen aufgrund von Gebührenordnungen und vereinbarten Pauschalen abrechnen, haben die Ausgaben der Sozialversicherer häufig wenig mit den tatsächlichen Ressourcenaufwendungen zu tun. Z.B. erhält die Kassenärztliche Vereinigung von den Krankenkassen eine Pauschale für die ambulante ärztliche Versorgung der Versicherten in einer Region in einem Jahr. Somit stellen in dieser Perspektive zusätzliche oder eingesparte Arztbesuche keine Kosten oder Kostenersparnisse dar. Auch werden bei der Sozialversicherungsperspektive Selbstbeteiligungen der Patientinnen und Patienten oder Leistungen der Arbeitgeber nicht als Kosten erfasst.

Tabelle 1 gibt einen internationalen Überblick über die in einzelnen Ländern verwendete Perspektive bei der Evaluation von Arzneimitteln. Zudem gibt die Tabelle an, ob in den genannten Ländern Schwellenwerte für Kosten verwendet werden. Wenn z.B. der Schwellenwert 30.000 \$ pro Nutzeinheit ist, dann werden alle Leistungen, die teurer sind, nicht oder nicht voll vom Gesundheitsversorgungssystem finanziert. In Deutschland wird weder eine volkswirtschaftliche oder gesellschaftliche Perspektive bei der Evaluation im Gesundheitswesen eingenommen noch hat man sich zu einem Schwellenwert durchringen können. Stattdessen findet die Bewertung der Kosten (und der Kostenersparnisse) aus Sicht der Krankenkassen statt. Das bedeutet, dass z.B. Kosten (und Kostenersparnisse) der Arbeitgeber oder Selbstbeteiligungen der Patientinnen und Patienten bei der Bewertung von Arzneimitteln unberücksichtigt bleiben. Ohne Schwellenwert bleibt die KNA ein „zahnloser Tiger“, denn wie soll man dann begründen, dass ein Arzneimittel zu teuer ist.

Tabelle 1: Perspektive der Kostenbewertung in verschiedenen Ländern

	Gesellschaftliche Perspektive	Perspektive der Zahlungsträger
Orientierung an einem Schwellenwert	Australien, Kanada	England, Neuseeland, Schottland
Kein Schwellenwert	Finnland, Niederlande, Frankreich	Deutschland

In der deutschen Diskussion gibt es durchaus auch Stimmen, die bezweifeln, ob die Kosten im Gesundheitswesen überhaupt entscheidungsrelevant sein sollen, da ja Gesundheit das „höchste Gut“ ist. Es wird eine „Ökonomisierung“ der Medizin befürchtet.

Es sind zwei Situationen denkbar, in denen für kollektive Entscheidungsträger im Gesundheitssystem die Ergebnisse einer KNA und die darin ermittelten Kosten hilfreich sein könnten:

1. Die angebotenen indizierten und medizinisch sinnvollen Leistungen werden zu Preisen angeboten, die aufgrund einer Kosten- und Nutzenabwägung als zu hoch anzusehen sind. Es gilt, Erstattungspreise so festzulegen beziehungsweise mit den Leistungsbrin-

gern Preise so zu verhandeln, dass das Kosten-Nutzen-Verhältnis akzeptabel ist und die Leistungen weiter angeboten werden.⁵

In diesem Falle ist die KNA ein Instrument der Preisregulierung. Solange die Gesundheitsleistungsanbieter diese akzeptieren und sich nicht aufgrund zu niedriger Preise vom Markt zurückziehen, entsteht weder eine Verteilungsproblematik (außer der Schlechterstellung der Anbieter) noch ein wirklicher Anlass, einen methodischen Streit zu führen.

2. Es reichen die zur Verfügung stehenden Mittel nicht für alle Ausgaben für indizierte und medizinisch sinnvolle Leistungen zur Versorgung der Patientinnen und Patienten aus.

In diesem Fall ist die Frage zu beantworten, wie entweder das System – zum Beispiel durch Selbstbeteiligungen, Zusatzbeiträge und eine zusätzliche Steuerfinanzierung – mit zusätzlichen Mitteln ausgestattet werden kann oder es sind Leistungen aus dem Leistungskatalog zu streichen. Letzteres wird in Deutschland von allen politischen Parteien und Entscheidungsträgern (derzeit noch) abgelehnt. Das vom früheren Ärztekammerpräsidenten Jörg-Dietrich Hoppe angestoßene Thema Priorisierung wurde gleich tabuisiert. Eine „heimliche Rationierung“ (Hoppe nach DÄB 27.5.2011) wird einer expliziten Diskussion, wie mit der Mittelknappheit umgegangen werden soll, vorgezogen.

Dabei sind nach höchstrichterlicher Rechtsprechung Wirtschaftlichkeitsüberlegungen aus verfassungsrechtlicher Sicht zulässig und angesichts der großen sozialpolitischen Herausforderungen auch legitim. Das Bundesverfassungsgericht führt in dem sogenannten Nikolausurteil (6.12.2005) aus:

„Der Leistungskatalog der gesetzlichen Krankenversicherung darf auch von finanzwirtschaftlichen Erwägungen mitbestimmt sein. Gerade im Gesundheitswesen hat der Kostenaspekt für gesetzgeberische Entscheidungen erhebliches Gewicht.“⁶

5 BUNDESMINISTERIUM FÜR GESUNDHEIT 2011.

6 BVerfG, 1 BvR 347/98, Rn. 58.

Allerdings schränkt das Bundesverfassungsgericht diese generelle Feststellung wie folgt ein:

„Es ist mit Art. 2 Abs. 1 GG ... nicht vereinbar, den Einzelnen ... einer Versicherungspflicht in der gesetzlichen Krankenversicherung zu unterwerfen und für seine an der wirtschaftlichen Leistungsfähigkeit ausgerichteten Beiträge die notwendige Krankheitsbehandlung gesetzlich zuzusagen, ihn andererseits aber, wenn er an einer lebensbedrohlichen oder sogar regelmäßig tödlichen Erkrankung leidet, für die schulmedizinische Behandlungsmethoden nicht vorliegen, von der Leistung einer bestimmten Behandlungsmethode durch die Krankenkasse auszuschließen und ihn auf eine Finanzierung der Behandlung außerhalb der gesetzlichen Krankenversicherung zu verweisen. Dabei muss allerdings die vom Versicherten gewählte andere Behandlungsmethode eine auf Indizien gestützte, nicht ganz fernliegende Aussicht auf Heilung oder wenigstens auf eine spürbare positive Einwirkung auf den Krankheitsverlauf versprechen.“⁷

Drei Jahre später wiederholt das Bundesverfassungsgericht seine Auffassung: „Es bedarf ... einer besonderen Rechtfertigung, wenn dem Versicherten Leistungen für die Behandlung einer Krankheit und insbesondere einer lebensbedrohlichen oder regelmäßig tödlich verlaufenden Erkrankung durch gesetzliche Bestimmungen oder durch deren fachgerichtliche Auslegung und Anwendung vorenthalten werden.“⁸

In die allgemeine Debatte um die KNA hat sich 2011 auch der Deutsche Ethikrat mit einer 90-seitigen Stellungnahme eingeschaltet.⁹ Der Stellungnahme folgt ein 30-seitiges Sondervotum der Ethikerin Wilma Lübbecke. Der Ethikrat kommt zum Ergebnis: „Eine sachliche Debatte erfordert ... die Einbeziehung medizinischer, ökonomischer, ethischer und juristischer Expertise... Verteilungsentscheidungen sind nicht allein Gegenstand wissenschaftlicher Expertise... Letztlich sind Entscheidungen

7 BVerfG, 1 BvR 347/98, Rn. 64.

8 BVerfG, 1 BvR 1665/07 vom 30.6.2008, Rn. 7.

9 DEUTSCHER ETHIKRAT 2011.

über den Umfang solidarisch finanzierter Leistungen ethische Entscheidungen, die im gesellschaftlichen Diskurs und auf politischem Wege getroffen werden müssen.“¹⁰ Der Forderung nach einem interdisziplinären und gesellschaftlichen Diskurs ist zuzustimmen. Wenn es sich „letztlich“ um ethische Entscheidungen handelt, würde man sich einen konkreten, operationalisierbaren Beitrag der Ethikerinnen und Ethiker wünschen.

3. Bewerten der Nutzen

Nach § 35b Abs.1 S.4 SGB V sollen bei der Bewertung des Patientennutzens insbesondere folgende Aspekte berücksichtigt werden:

- die Verbesserung des Gesundheitszustandes,
- die Verkürzung der Krankheitsdauer,
- die Verlängerung der Lebensdauer,
- die Verringerung der Nebenwirkungen und
- die Verbesserung der Lebensqualität.

Die Gesetzesformulierung macht zwei Dinge deutlich: Erstens gilt bei dem Nutzen medizinischer Leistungen die Patientenperspektive. Zweitens macht die Aufzählung deutlich, dass es sich beim Nutzen um ein mehrdimensionales Phänomen handelt. Hieraus resultieren nun zwei generelle methodische Aggregationsprobleme: das personelle Aggregationsproblem und das dimensionale Aggregationsproblem. Das personelle Aggregationsproblem beschreibt die Schwierigkeit, den Nutzen eines Patienten oder einer Patientin mit dem Nutzen eines anderen Patienten oder einer anderen Patientin zu aggregieren. Wenn z.B. bei der Verwendung zweier sich gegenseitig ausschließender Therapien (Operation oder konservative Behandlung) aus einer gegebenen Patientenpopulation bei der einen nur sechs und bei der anderen Therapie zwölf (andere) Personen versterben, kann man dann sagen, dass die eine Therapie doppelt so gut ist wie die andere? Dies kann man nur, wenn jede und jeder Verstorbene mit dem gleichen Gewicht bewertet wird. Lehnt man eine solche Aggregation ab (wie es z.B. der Deutsche Ethikrat anmerkt), so ist eine solche Aussage nicht möglich. Es ist so-

10 DEUTSCHER ETHIKRAT 2011, S. 95.

gar keine wertende Aussage möglich, denn bei der einen Therapie sind Menschen gestorben, die mit der anderen überlebt hätten et vice versa.

Beide Aggregations-Probleme – ohne dass sie so benannt worden sind – haben in Deutschland zu einer heftigen Diskussion geführt. Auch der Hinweis im Gesetz, dass Grundlage der KNA die „in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie“ sein sollen (§ 35b (1) Satz 5), konnte die Kontrahenten nicht beruhigen.

Gerade die evidenzbasierte Medizin und die gesundheitsökonomische Forschung haben für diese Aggregationsprobleme Standards gesetzt, die in zahllose Guidelines gegossen wurden: Bei der Analyse der Ergebnisse von klinischen Studien zählt jede Probandin bzw. jeder Proband statistisch gleich. Für die Aggregation der verschiedenen Dimensionen des Nutzens wurden ebenfalls von der Forschung Konzepte entwickelt, von denen der SF36 und der EQ 5D für die Aggregation der Dimensionen der Lebensqualität und das QALY-Konzept für die Aggregation von Lebensdauer und Lebensqualität die bekanntesten sind.

Statt dieser weitgehend im internationalen Bereich anerkannten Konzepte wollte man einen eigenen Weg finden, der letztlich dazu geführt hat, dass die KNA im deutschen Gesundheitswesen kaum Anwendung findet. Auch die Bewertungen von Arzneimittelinnovationen nach dem AMNOG-Verfahren bedienen sich nicht einer Abwägung der Kosten und der Nutzen. Es wird lediglich verlangt, den Budgeteffekt der Innovation für die Krankenkassen zu schätzen. Der Nutzen bleibt disaggregiert und man fokussiert sich nur auf die klinische Wirksamkeit.

Tabelle 2 fasst die Konsequenzen für die Nutzenbewertung zusammen. Ist man bereit, sich auf ein einheitliches Nutzenkonzept zu einigen und ist man bereit, das dimensionale Aggregationsproblem zu lösen, so gelangt man zu einer umfassenden indikationsübergreifenden Bewertung medizinischer Leistungen und Programme, wie es z.B. das englische National Institute for Health and Care Excellence (NICE) und das schwedische Pharmaceutical Benefits Board (Läkemedelsförmånsnämnden - LFN) vornehmen. Lehnt man Indikationen übergreifende Konzepte – z.B. mit dem Argument der Individualität und Unterschiedlichkeit der Patientinnen und Patienten - ab und verschließt man sich einer Zusammenfassung der Nutzendimensionen in einem Maß, dann bleibt nur die Lösung, wie sie vom deutschen IQWiG vorgeschlagen wurde.

Tabelle 2: Methoden der Nutzenerfassung

	Interindikations-Vergleich	Intraindikations-Vergleich
Multidimensionale Nutzenskala	z.B. QALY-Konzept	Indikationsspezifische Nutzen-Mess-Instrumente
Monodimensionale Nutzenskala	Kosten-Effektivitätsanalyse	Ansatz des deutschen IQWiG

Um die Komplexität und Gedankenansätze zu verdeutlichen, wird noch einmal spezifisch auf die deutsche Diskussion des QALY-Konzepts eingegangen. In der oben bereits erwähnten Stellungnahme des Deutschen Ethikrats geht dieser auf das QALY-Konzept ein. Da beim QALY-Konzept neben dem Lebensqualitätsgewinn aufgrund einer Therapie auch die zeitliche Dauer dieses Gewinns berücksichtigt wird, sieht der Ethikrat in der Verwendung des Konzeptes eine Diskriminierung älterer und chronisch kranker Patientinnen und Patienten.¹¹ Zu Recht weist der Deutsche Ethikrat darauf hin, dass das QALY-Konzept auf Werturteilen beruht. Allerdings führt er nicht aus, dass seine Aussage auch ein Werturteil beinhaltet: Es soll offenbar nicht jedes Lebensjahr, sondern jede Person gleich gewichtet werden. Wenn ein erfolgreich behandeltes Kind sein ganzes Leben von der Therapie eine Lebensqualitätssteigerung hat, dann ist das – nach Ansicht des Deutschen Ethikrates – gleich zu bewerten wie eine gleich hohe Lebensqualitätssteigerung einer Therapie, die wenige Tage vor dem Tode erfolgt.

Der Ethikrat wendet sich dagegen „... dass knappe medizinische Ressourcen QALY-maximierend eingesetzt werden sollen“.¹² „Bei der übergeordneten Anwendung der Ergebnisse geht es also nicht um die Erreichung der maximalen Zahl von QALYs für einen bestimmten Patienten, sondern – auf das Kollektiv bezogen – um die Optimierung der QALYs entsprechend einem zur Verfügung stehenden finanziellen Budget.“¹³ In dieser Aussage steckt mindestens ein Missverständnis und ein Widerspruch. In keinem Land der Welt werden ökonomische Evaluationsstudien zur Maximierung von QALYs (der Begriff kommt mehrfach in der Stellungnahme vor) eingesetzt. Vielmehr geht es darum,

11 DEUTSCHER ETHIKRAT 2011.

12 DEUTSCHER ETHIKRAT 2011, S. 63.

13 DEUTSCHER ETHIKRAT 2011, S. 62.

Gesundheitsleistungen aus dem Leistungskatalog zu eliminieren oder deren Preise zu drücken, deren Kosten angesichts des erwarteten Nutzens für die Patientinnen und Patienten zu hoch sind. Ein festgelegter Schwellenwert, wie beim viel zitierten englischen NICE (z.B. 30.000 BP/QALY), der von den Entscheidungstragenden flexibel gehandhabt wird, gibt einen Anhaltspunkt, welchen Mindeststandard eine Gesundheitsleistung bringen muss, damit sie von der Gemeinschaft finanziert wird. Ebenso verfährt die evidenzbasierte Medizin, indem sie Mindeststandards festlegt: Eine Therapie, die von 1.000 Patientinnen und Patienten 500 nutzt, für 100 keinen Effekt zeigt und 300 mit ihren Nebenwirkungen schadet, mag den Mindeststandard beim medizinischen Nutzen-Nebenwirkungsverhältnis nicht erfüllen; eine Therapie, die 800 Patientinnen und Patienten nutzt und nur 50 schadet, erfüllt ggf. den Mindeststandard. Übrigens zeigt das Beispiel, dass man bei der Auswertung klinischer Studien und der Festlegung von Mindeststandards kaum um eine utilitaristische Werthaltung (Aggregation individueller Nutzenwerte) herumkommt; es sein denn, eine Therapie ist dominant überlegen.

Da es „ohne Knappheit keine Verteilungsprobleme“ gibt (S.13) und ohne Knappheit keine Notwendigkeit für KNA besteht, wird man Knappheitsprobleme nur lösen können, wenn man Werturteile über Verteilungsfragen fällt.¹⁴ Das QALY-Konzept basiert auf dem Werturteil, dass jedes Lebensjahr eines Patientenkollektivs mit dem gleichen Gewicht in die Nutzenbewertung eingeht. Wer andere Wertvorstellungen hat, sollte diese explizit formulieren, wenn er einen nützlichen Beitrag zur Lösung des Ressourcenallokationsproblems leisten möchte.

Die Positionierung des Deutschen Ethikrats und der Wunsch des IQWiG, mit dem Konstrukt der Effizienzgrenze (das hier nicht weiter erläutert wird und auch praktisch keine Rolle spielt) einen methodischen Sonderweg zu gehen,¹⁵ hat dazu geführt, dass im Gegensatz zum Ausland die KNA im deutschen Gesundheitssystem eine marginale Rolle spielen. Explizit sind KNA lediglich bei Arzneimitteln vorgesehen (§ 35b SGB V). Sie kommen allerdings nur zur Anwendung, wenn

14 SCHULENBURG 2010.

15 GREINER, KUHLMANN, SCHWARZBACH 2010.

der Hersteller dies verlangt, oder wenn nach gescheiterten Preisverhandlungen zwischen Arzneimittelhersteller und dem Spitzenverband Bund der Krankenkassen der Schiedsspruch über die Preisfestsetzung vom Pharmaunternehmen nicht akzeptiert wird. Beide Fälle kommen praktisch nicht vor.

4. Lehren für andere Politikbereiche

Aus den Ansätzen der Evaluation im Gesundheitswesen lernt man Folgendes: Die Evaluationsforschung hat leistungsfähige Konzepte entwickelt, um die Kosten und Nutzen von Entscheidungen zu evaluieren und alternative Entscheidungen bezüglich ihrer Kosten-Effektivität zu bewerten. Allerdings basiert jedes Bewertungskonzept auf Werturteilen bzw. Paradigmen. Man kommt nicht darum herum, diese Paradigmen zu diskutieren und bei gesellschaftlichen Entscheidungen einen Konsens über die zu verwendenden Werturteile herbeizuführen. Im Gesundheitswesen hat man es sich hiermit in Deutschland besonders schwierig gemacht. Dies liegt auch daran, dass es in der medizinischen Versorgung um menschliches Leid und Tod geht. Wir tun uns schwer, menschliches Leid und Tod zu bewerten; obwohl dies nötig ist. Denn findet eine offene transparente und nachvollziehbare Bewertung der Kosten und Nutzen nicht statt, so findet sie versteckt, intransparent und damit auch willkürlich statt.

Literatur

Dasgupta, A.K., Pearce, D.W. (1972). „Cost-Benefit Analysis, Theory and Practice“. London: Macmillan.

Little, I.M.D., Mirrless, J. (1969). „Manual of Industrial Project Analysis in Developing Countries“. Vol 2. Social Cost Benefit Analysis. Paris: OECD.

Deutscher Ethikrat (Hrsg.) (2011). „Nutzen und Kosten im Gesundheitswesen – Zur normativen Funktion ihrer Bedeutung: Stellungnahme“. Berlin.

Schöffski, O., Schulenburg, J.M. G.v.d. (Hrsg.) (2011). „Gesundheitsökonomische Evaluationen“. Berlin: Springer.

Schulenburg, J.M. G.v.d. (2010). „Rationierung und Gerechtigkeit“. Gesundheitsökonomie & Qualitätsmanagement 4, 158-161.

Greiner, W., Kuhlmann, A., Schwarzbach C. (2010). „Ökonomische Beurteilung des Effizienzgrenzenkonzeptes“. Gesundheitsökonomie & Qualitätsmanagement 15, 241–250.

Bundesministerium für Gesundheit (2011). „Faire Preise für Arzneimittel. Die Preisgestaltung in der GKV nach dem Arzneimittelmarktneuordnungsgesetz“.

5. Politikfeld Familienpolitik

Evidenzbasierung in der Familienpolitik: Wie sie möglich wurde, was sie ausmacht und warum sie Promotoren aus der Gesellschaft braucht

Von MALTE RISTAU

Korrespondenzadresse:

Prognos AG

Goethestr. 85

10623 Berlin

winklerristau@hotmail.com

1. Politikwechsel

Über Jahrzehnte hatte in der Bundesrepublik eine mit hohem finanziellen Aufwand betriebene Familienpolitik unbefriedigende Ergebnisse erzielt. Sie verstand sich im Wesentlichen als soziale Ausgleichspolitik über Transfers (Lastenausgleich); wurde der Mitteleinsatz erhöht, galt das weithin als gute Politik. Dass andere Ergebnisse nicht mehr finanziellen Aufwand erforderten, belegte 2002 eine Studie der Bertelsmann-Stiftung. Anderswo waren die Kinder besser gefördert, den Müttern wurde Erwerbstätigkeit stärker ermöglicht; Armutsrisiken waren geringer und Väter in den Familien aktiver. Außerdem wurden in den Ländern, die eine dezidiert Vereinbarkeit fördernde Politik umsetzten, mehr Kinder geboren. Unerwartet, aber nicht zufällig gab der Bundeskanzler 2002 eine familienpolitische Regierungserklärung mit neuen Akzenten ab.¹ Schröders Sicht auf das von ihm vorher so titulierte „Gedöns“ hatte sich aufgrund von Wahlanalysen geändert, von *Anpassungslernen* ließe sich insofern sprechen. Gute Vereinbarkeit war der von Schröder umworbenen *Neuen Mitte* wichtig und die Kompe-

1 MACKROTH, RISTAU 2002.

tenz dafür wurde der SPD quasi als Vorschuss zugeschrieben, weil die Partei als moderner galt.

Schröder vollzog einen Führungswechsel im Familienressort (von Christine Bergmann zu Renate Schmidt) und im Bundesfamilienministerium (BMFSFJ) wurde ab Anfang 2003 top down ein Politikwechsel vorbereitet. Voraussetzung dafür war ein planmäßiges Changemanagement mit Änderungen, die vor allem produktive Arbeitsstrukturen unabhängig von Geschäftsverteilungsplänen und das Heranziehen anderer externer Kompetenzträger umfassten. Das Wissen über Wirkungen von Leistungen war zu diesem Zeitpunkt gering, eine Übersicht der Bundesregierung über die wesentlichen Leistungen gab es nicht und ökonomische Betrachtungen waren in diesem Politikfeld nicht üblich. Als erstes wurde der im Ministerium erarbeitete Auftrag für den anstehenden 7. Familienbericht neu formuliert. Internationale Erfahrungen mit unterschiedlichen Politikpfaden sollten geprüft und mit Bezug auf ausgewählte Ziele wirkungsstarke Instrumente erörtert werden. Auf Grundlage einer systematischen Strategie verstand sich die neue Hausspitze deutlich abgehoben von Haltungen zuvor als gestaltende Instanz eines eigenständigen Politikfeldes und nicht mehr als mitredend bei einer Querschnittsaufgabe. Die internen Lernprozesse waren nicht einfach und erwiesen sich später als fragil, weil stark abhängig von Einzelpersonen, Hausleitungen und Führungskräften der oberen Ebene.

Die Neuorientierung nach außen mittels Leadership erfolgte gegen latente Zweifel und offene Widerstände in Fachverbänden, Parteien und Fraktionen. Jedoch überzeugte viele in Medien und Öffentlichkeit die Übereinstimmung mit den mehrheitlichen Lebenswünschen der Bevölkerung wie wesentlichen Empfehlungen aus der laufenden Arbeit der Kommission zum Familienbericht. Diese war Mitte 2003 unter Vorsitz von Hans Bertram eingerichtet worden und arbeitete zwei Jahre.² Als Ergebnis eines *komplexen Lernprozesses* von Politik und Wissenschaft wurden ab 2004 als neuer Weg monetäre Förderung mit Elterngeld und Kinderzuschlag (KiZ) wirkungsorientiert ausgerichtet, nach einer Milliardenhilfe an die Länder für Ganztagschulen eine am Bedarf orientierte Betreuung für die unter Dreijährigen bedarfsgerecht ausgebaut sowie

2 Mitglieder der Kommission waren u.a. die Familien- und Bildungsökonomin C. Katharina Spieß sowie die seinerzeitige Direktorin des IAB, Jutta Allmendinger.

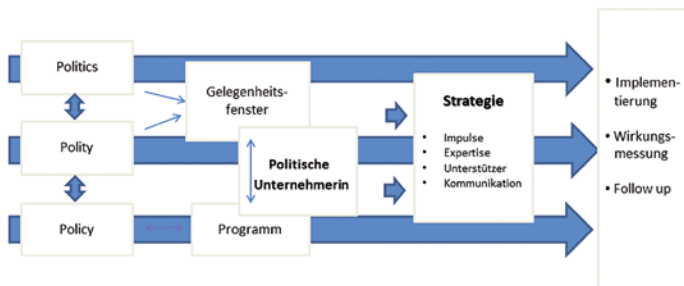
die Arbeitgeber ermuntert und mit Programmen unterstützt, familienbewusster zu handeln.

Die Leitidee des so ausgefüllten Drei-Dimensionen-Mix prägte für einige Jahre die öffentliche Debatte.³ Der in der Öffentlichkeit wie in der Familienforschung als Paradigmenwechsel bezeichnete neue Akzent lag darin, dass Familienpolitik deutlich stärker als zuvor als auf Nachhaltigkeit angelegte Investition in Humanvermögen verstanden wurde.⁴ Erwerbstätigkeit von Müttern sollte den Lebensplänen entsprechend ermöglicht werden und Kinder aus allen Schichten sollten besser gefördert werden. Erfahrungen aus anderen Ländern, insbesondere aus Nordeuropa, wurden berücksichtigt, namentlich beim Elterngeld, eine Art Lohnersatzleistung, sowie beim Ausbau der Kinderbetreuung für die unter Dreijährigen mit dem Tagesbetreuungsausbaugesetz (TAG, 2005) sowie dem darauf aufbauenden Kinderfördergesetz (KiFöG, 2008).

2. Strategie

Eine Ex-post-Politikfeldanalyse ergibt in der Summe ein lehrbuchhaftes strategisches Vorgehen (Abb.1).

Abbildung 1: Modifizierte Politikfeldanalyse



Für eine wissenschaftlich basierte Reformstrategie wurden günstige Zeitfenster zum aktiven Agenda-Setting genutzt (*Politics*). Zwei offen-

3 Zeit, Infrastruktur und Geld lauteten die drei Dimensionen, die im 7. Familienbericht ausgewiesen und für eine gestaltende Umsetzung ausdrücklich als zusammenhängend apostrophiert wurden.

4 RISTAU 2005.

sive Ministerinnen, Renate Schmidt und Ursula von der Leyen, ergriffen als politische Unternehmerinnen Chancen jenseits der Koalitionsverträge (*Polity*). Indem sie eine breite gesellschaftliche Unterstützung für ihr Programm mobilisierten, haben sie nacheinander ihre Parteien, SPD und CDU, mitgezogen: gegen zum Teil heftigen Widerstand, in die gleiche Richtung, was Werte, Ziele und Instrumente betraf (*Policy*). Eine von den Ministerinnen aktiv betriebene Advocacy-Koalition mit operativem Unterbau verschärfte den Erfolgsdruck auf Parlament und Regierung. Ab 2004 wurden mit Unterstützung durch ein Servicebüro des BMFSFJ mehrere hundert Lokale Bündnisse für Familien gegründet sowie ein Jahr später ein beim DIHK angesiedeltes Unternehmensnetzwerk Erfolgsfaktor Familie eingerichtet, an dem sich nach wenigen Jahren rund 7000 Betriebe auf Dauer beteiligten. Beide Netzwerke wirkten als Kommunikationskanäle und dienten dezentral zur Mobilisierung für die intendierten Veränderungen.

Mit Stakeholdern aus Wirtschaft und Gesellschaft war 2004 eine informelle Verständigung auf programmatische Eckpunkte einer *nachhaltigen Familienpolitik* erfolgt, die unter dem Rubrum *Allianz für die Familie* gemeinsam vertreten und von einigen Akteuren arbeitsteilig verfolgt wurden.⁵ Entstanden waren die Eckpunkte im Dialog von Führungskräften des BMFSFJ mit Sachverständigen, die ihren Transfer wissenschaftlicher Befunde als „pragmatische Politikberatung“ verstanden.⁶ Für die Medien reizvoll und für die Bevölkerung anziehend erwies sich die bis dahin ungewöhnliche Rezeptur der Argumentation, die stetig und immer unterlegt mit neuen Daten vorgetragen wurde. Je stärker evidenzbasiert von den Ministerinnen, ihren Beratern und Unterstützern argumentiert wurde, desto mehr stieg die Zustimmung in der Bevölkerung.

Die Argumentation beschränkte sich aber nicht auf einschlägige Daten, sondern sie beruhte auf einer Kombination von kommunitaristischer Verantwortung und ökonomischer Rationalität; aus modernen Rollenbildern und progressiver Wertschätzung der „alten“ Lebensform

5 SCHMIDT, MOHN 2004. Zu den Unterstützern zählten u.a. der Präsident des DIHK, Ludwig Georg Braun sowie der Ratsvorsitzender der EKD, Wolfgang Huber.

6 Zu dieser Haltung vgl. begründend RÜRUP 2009.

Familie, die als soziale Mitte der Gesellschaft verstanden wurde.⁷ Bezogen auf Wirkungsorientierung bedeutete das die Koppelung von politischen Gestaltungszielen, zum Beispiel Vereinbarkeit, konkretisierten Werten mit an Kennziffern ausgewiesener Evidenz. Besondere Aufmerksamkeit erreichten die beiden Ministerinnen immer dann, wenn sie seitens der Politik bislang ungewohnte Argumente zur betrieblichen Rendite von Vereinbarkeit, zu volkswirtschaftlichen Wachstumseffekten von Familienfreundlichkeit oder zu nicht realisierten Kinderwünschen und ihrer Bedeutung für die demografische Entwicklung in die Debatte einbrachten. Zum Kurswechsel im BMFSFJ zählte denn auch, neben dem Heranziehen von internationalen Benchmarks, eine intensive Kommunikation von Ergebnissen ökonomischer Studien.

Der Anfang wurde gleich im Januar 2003 gemacht, als C. Katharina Spieß (DIW) erstmals einen volkswirtschaftlichen Nutzen des Ausbaus von Kinderbetreuung belegte. Im November 2003 erläuterte der Wirtschaftsweisen Bert Rürup an der Seite der Familienministerin den ökonomischen Charme des Elterngeldes und fand dafür viel Beachtung. Die Prognos AG wies 2004 in drei Studien nach, dass und wie konsequente Familienfreundlichkeit für Unternehmen unterschiedlicher Größen und Branchen Rendite bringt. In Kooperation mit der *ZEIT* wurde 2004 und 2008 die Familienfreundlichkeit aller Kreise und Städte anhand aussagekräftiger Kennziffern überprüft. Die Stärken und Schwächen offenlegenden Ergebnisse wurden als „Familienatlas“ mit fulminantem Echo veröffentlicht und gemeinsam mit insbesondere dem DIHK auch dezentral antreibend kommuniziert. Das IW Köln stufte 2006 die nachhaltige Familienpolitik als interessanten Wachstumspfad ein und betonte kausale Wirkungszusammenhänge zu den zentralen wachstumspolitischen Parametern Personen, Arbeitszeit und Produktivität. Weitere ökonomische Studien namhafter Personen und Institutionen sind seither in dieser Ausrichtung gefolgt.

Die *nachhaltige Familienpolitik* mit der Betonung von empirischer Evidenz als einem bedeutsamen Merkmal entsprach dem wirkungsorientierten Ansinnen der Agenda 2010, Menschen zu befähigen und Anreize zur Erwerbsarbeit zu stärken.⁸ Wichtige Projekte, TAG und KiZ,

7 MOHN, VON DER LEYEN 2007.

8 RISTAU 2006.

wurden als Teil der Agenda 2010 im Gesetzgebungsverfahren und im Finanztableau behandelt. Die programmatische Leitlinie und anschließende Erfolge durch Wirksamkeit veranlassten drei Arbeitsminister und einen Finanzminister 2004 bis 2009, den Kurs der beiden Familienministerinnen zu unterstützen, den diese mit zum Teil plebiszitären Anklängen verfolgten. Der gar nicht hoch genug zu bewertende Umstand einer Kontinuität für kommunikationsstarke Strategie und inhaltliches Konzept über eine Legislatur hinaus, ungeachtet eines politischen Farbenwechsels 2005 an der Spitze des BMFSFJ, war ein wesentlicher Garant für die sich recht bald einstellenden tatsächlichen Erfolge bei den Ergebnissen.⁹ Das dargestellte Politikmuster konnte immerhin einige Jahre als Beispiel für „gutes Regieren“ erfolgreich praktiziert werden.

3. Wirkungsorientierung

Neu eingeführte oder veränderte Leistungen wurden sämtlich von einem Monitoring begleitet. Vom Ministerium wurde erstmals 2006 in Gestalt eines Finanztableaus eine verlässliche Gesamtschau über Zahl und Volumina der familien- und ehebezogenen Leistungen erstellt. Die wissenschaftliche Politikberatung wurde 2006 in einem Kompetenzzentrum weiter professionalisiert; eine Agenda-Gruppe von Sozial- und Wirtschaftswissenschaftlerinnen und –wissenschaftlern wurde als ungewöhnliche Struktur von der Ministerin persönlich geleitet und traf sich regelmäßig. Aus dieser Gruppe gab es eine Reihe von Anregungen namentlich zu einer veränderten Staffelung des Kindergeldes, zur effektiveren Unterstützung von Alleinerziehenden, zur Optimierung des Kinderzuschlags sowie 2009 schließlich ein Memorandum zur Zeitpolitik. Seit 2008 wurde jährlich in einem Familienreport die Entwicklung der Familien in Deutschland dargestellt und Rechenschaft über den Grad der Zielerreichung bei wesentlichen Leistungen mit Erläuterung von Kausalitäten abgelegt.

Das Ansehen der offensiv begründeten und neu ausgerichteten Familienpolitik in der Bevölkerung verbesserte sich stetig und erreichte ein in der Politik ungewöhnliches Niveau. Gleiches galt für die Bewer-

9 Die großen Fortschritte bei Müttererwerbstätigkeit und Väterengagement, Kitaplätzen, wirtschaftlicher Stabilität auch im unteren Einkommensbereich sowie Geburtenrate sind mittlerweile unstrittig; vgl. die Daten im jüngsten Familienreport 2017, hrsg. vom BMFSFJ.

tung der Ministerinnen Schmidt und von der Leyen, die etwa im Politbarometer oder im Deutschlandtrend erstmals und bis auf weiteres letztmals für dieses Politikfeld als Spitzenkräfte der Bundespolitik positiv wahrgenommen wurden. Über die Familienpolitik hatten sich laut Allensbach vor 2003 lediglich unter 30% der repräsentativ Befragten positiv geäußert, demgegenüber befanden zwischen 2007 und 2009 fast 60%, dass „die Bundesregierung bei der Familienförderung gute Arbeit leiste“. Nach vier Jahren evidenzbasierter Reformpolitik vertrat eine deutliche Mehrheit der Bevölkerung die Auffassung, dass sich seit 2005 „die Bedingungen für Familien mit Kindern spürbar verbessert haben“.¹⁰ Die Medien registrierten das dauerhafte Meinungsklima und würdigten die empirisch gestützte Erfolgsgeschichte. Laut maßgeblichem sozial- und wirtschaftswissenschaftlichen Urteil entwickelten sich die neuen Leistungen nämlich so wirksam und nachhaltig wie intendiert.¹¹ Das ist aber ein eigenes lohnendes Thema im Rahmen der Erfolgsgeschichte der Bundesrepublik Deutschland.

In gewisser Weise folgerichtig schloss in diesem Politikfeld ähnlich wie kurz zuvor in der Arbeitsmarktpolitik eine weitere systematische Maßnahme eine bis dahin politisch wie wissenschaftlich nicht nur in Deutschland ignorierte Lücke. Nach einem sorgfältigen Vorlauf startete noch vor der Bundestagswahl 2009 die interdisziplinär angelegte „Gesamtevaluation der ehe- und familienbezogenen Leistungen“. An öffentlich nicht nur proklamierten, sondern sorgfältig erklärten Zielen sollte ausgewiesene Wirksamkeit zum politischen Maßstab für Leistungen werden. Angelehnt an den 7. Familienbericht hatten sich BMFSFJ und Bundesfinanzministerium (BMF) auf fünf zentrale Ziele verständigt, deren Erreichen überprüft werden sollte: Wirtschaftliche Stabilität von Familien, Vereinbarkeit von Familie und Beruf, Nachteilsausgleich zwischen und für Familien, frühe Förderung und Wohlergehen von Kindern sowie Erfüllung von Kinderwünschen. Anhand der 13 größten Leistungen analysierten wissenschaftliche Teams bis 2013 mit unterschiedlichen Methoden den Grad der Zielerreichung (Effektivität), aber auch die Relation (Effizienz) von Zielerreichung zu fiskalischem Aufwand. Mit „Akzeptanz-Analysen“ wurden Bekanntheit und Bewertung der wich-

10 ALLENSBACH 2009.

11 BONIN 2014.

tigsten Leistungen in der Bevölkerung wie bei den Antragsberechtigten sowie die Muster ihrer Inanspruchnahme untersucht.

Das von Prognos organisierte und moderierte Projekt erwies sich als fruchtbar; es gab allerdings auf beiden Seiten, Wissenschaft und Politik, kulturelle Fremdheit und eingeschränkte Beratungsfähigkeit. Recht unterschiedlich waren Auffassungen nicht nur über Tempi und Verständlichkeit. Vieles, was Wissenschaftlerinnen und Wissenschaftler hochinteressant finden, hat wenig oder keinen Gebrauchswert für Politikgestaltung. Zielkonflikte zwischen BMFSFJ und BMF blieben bestehen. So kritisiert das BMF ungeachtet gemeinsam betriebener und publizierter Evidenz-Analysen regelmäßig, dass die Ausgaben für das Elterngeld wachsen. Das BMFSFJ seinerseits stellt dagegen, dass politisch gewollte Effekte, insbesondere eine höhere Väterbeteiligung bei der Elternzeit, höhere Kosten hervorrufen.

Andererseits verschafften der Grad der erreichten Rationalität evidenzbasierter Politik und das zwischenzeitlich erfolgte Lernen über starke Refinanzierungseffekte familienbezogener Leistungen eine höhere Akzeptanz für Positionen des BMFSFJ unter dem Blickwinkel der Tragfähigkeit öffentlicher Finanzen. Angesicht der Wirtschafts- und Finanzkrise 2008-9 schließlich wurden als Teil des Rettungspaketes Leistungen für Familien sogar offensiv ausgebaut, weil sie vom BMF als stabilisierend begriffen wurden. Ergebnisse der Gesamtevaluation jedenfalls sollten, so das Bundeskabinett 2009, „fortlaufend“ in das Regierungshandeln eingespeist werden. Diese Absicht war Konsequenz aus der Erkenntnis, dass Wirkung Zeit braucht und Wirksamkeit Kontinuität erfordert.

4. Konjunkturen¹²

Vorbereitet worden war im Familienressort und erwartet in der Öffentlichkeit, so Auswertungen von Demoskopie und Medientenor, eine nächste Phase nachhaltiger Familienpolitik, die den Höhepunkt ihres Ansehens erreicht hatte. Gleichwohl erwies sich die Bindekraft des Geschaffenen trotz Kontinuität im Kanzleramt als zu schwach. Der Hochkonjunktur folgte eine tiefe Delle. Den tonangebenden Kräften der neuen Bundesregierung respektive Koalitionspartner ab 2009 war ein wirkungsorientiertes Profil der Familienpolitik ausweislich der Bestim-

12 RISTAU 2015.

mungen des Koalitionsvertrages gleichgültig. Man ist angesichts der vorliegenden Erfahrungen versucht, in diesem Falle von *pathologischem Lernen* zu sprechen. Ein fortgesetztes Profil der Familienpolitik wurde von Medien oder relevanten Akteuren allerdings auch nicht eingefordert. Die zuvor beteiligten Wissenschaftlerinnen und Wissenschaftler meldeten sich zu diesem Zeitpunkt ebenfalls nicht erkennbar zu Wort.

Unter gänzlich anderen Zeitumständen entstandene Leistungen wurden nachfolgend in klassischer Weise erhöht. Das laut Evaluation wirkungsstarke Elterngeld wurde gekürzt und das gegenläufige Betreuungsgeld, 2014 vom Bundesverfassungsgericht aufgehoben, unter Ministerin Kristina Schröder 2012 eingeführt. Als Begründung dafür wurde nicht Wirksamkeit ins Feld geführt, sondern eine spezifische Vorstellung von Gerechtigkeit: Wenn die einen etwas bekommen, müssen die anderen auch etwas erhalten. Für die Familienpolitik war es am Ende mangels Gestaltungsehrgeizes nach zeitgemäßen Zielen eine verlorene Legislatur. Gegen den Verlust an Modernität und Evidenz erhoben schließlich 24 prominente Wissenschaftlerinnen und Wissenschaftler verschiedener Disziplinen aus Hochschulen und Instituten in einem außergewöhnlichen Aufruf in der ZEIT ebenso ihre Stimme wie vier Vorgängerinnen der Familienministerin, je zwei aus der SPD und der CDU.

Die Bevölkerung reagierte in Umfragen sensibel, stufte die Familienministerin in die zweite Liga der Bedeutung und der Zustimmung zurück. Die Zustimmung zur Familienpolitik sank innerhalb kurzer Zeit drastisch von 59 auf 34%. Ungeachtet dessen wurden aber die sich ansammelnden Ergebnisse der Gesamtevaluation regelrecht versteckt und am Ende in einer Veröffentlichung des BMFSFJ aus Sicht der federführenden Wissenschaftlerinnen und Wissenschaftler geradezu verdreht. Dagegen verwahrten sich in einer öffentlichen Stellungnahme die Präsidenten von DIW, ifo und ZEW; diese Institute veröffentlichten schließlich ihrerseits eine eigene Quintessenz zentraler Resultate der Gesamtevaluation.¹³ Dieser beispiellose Sachverhalt blieb wohl auch deshalb einigermaßen unbekannt, weil sich dazu Spitzenorganisationen der deutschen Wissenschaft nicht äußerten.

Der Koalitionsvertrag 2013 fiel aus Sicht einer evidenzbasierten Perspektive nicht überzeugender als sein Vorgänger aus. Allerdings stellte

13 BONIN 2013.

die neue Ministerin Manuela Schwesig die Ergebnisse der Gesamtevaluation 2014 ungeschminkt auf einer Pressekonferenz vor und machte sich den Ansatz zu eigen, der mittlerweile auch von der OECD mehrfach gewürdigt worden war.¹⁴ Dem ökonomischen Stellenwert von Familienpolitik wurde 2014-17 von Seiten des BMFSFJ über Gutachten und Fachgespräche wieder verstärkt Aufmerksamkeit gewidmet, um eine erfolgreiche Wirkungsorientierung zu ermöglichen. Mit einigen gesetzlichen Maßnahmen, Beispiel ElterngeldPlus, knüpfte die Ministerin im Verlauf der Legislatur an Empfehlungen aus der Gesamtevaluation an. Die Bevölkerung honorierte die Rückkehr zum Kurs vor 2010 mit wachsender Zustimmung, 51% waren es schließlich 2017 wieder, die sich positiv äußerten. Schwesigs Projekt einer Familienarbeitszeit, von namhaften Mitgliedern der Evaluationsgruppe unterstützt, fand hohe Aufmerksamkeit, wurde aber schließlich von ihrer Partei nicht zum Essential für eine weitere Zusammenarbeit in der nächsten Bundesregierung erklärt.

Mit einem *Zukunftsreport Familie 2030* wurde immerhin 2016 vom BMFSFJ anhand von Szenarien ein neuer Anlauf unternommen, die nachhaltige Familienpolitik stringent weiter zu entwickeln. Laufende Trends der Entwicklung von Familienleben werden darin ebenso fortgeschrieben wie optionale Politikmuster. Das deutsche BIP könnte, so das Chancenszenario, demnach 2030 um rund 70 Mrd. Euro höher liegen als im Falle eines „Weiter so“. Was sich zu einem Qualitätssprung hinsichtlich evidenzbasierter Politikgestaltung im Kontext eines neuen Verständnisses von inklusivem Wachstum hätte entwickeln können, wurde bislang nicht weiterverfolgt. Dazu dürfte ein erneuter Leitungswechsel im BMFSFJ im Frühjahr 2017 beigetragen haben. Ein (Wahl-) Jahr darauf lesen sich die Vereinbarungen für die Arbeit der neuen Bundesregierung wiederum unentschieden. Für ein fundiertes Fazit ist es im Sommer 2018 noch zu früh. Positive Überraschungen unter der Verantwortung der mittlerweile siebten Ministerin in 16 Jahren, Franziska Giffey, sind nicht ausgeschlossen.

5. Fazit

Eine nachhaltige Familienpolitik auf Grundlage von Evidenzen wurde über einige Jahre konsequent und mit ermutigenden Ergebnissen

14 PROGNOSES 2014.

praktiziert. Insofern lieferte die Politik der Bundesregierung ebenso wie im Feld Arbeitsmarkt zumindest temporär anders als häufig unterstellt gute Leistung.

Vernunft setzt sich freilich auch in diesem Falle nicht selbstlaufend durch; es gibt keinen linearen Fortschritt für evidenzbasierte Politik. In allen Feldern bleibt die Politik deshalb hinter ihren Möglichkeiten zurück und verspielt so Chancen für eine stärkere Legitimation unseres politischen Systems in unruhigen Zeiten.

Neigungen zu evidenzbasiertem Handeln der Mitarbeiterinnen und Mitarbeiter in Bundesministerien nehmen selbst dann ab, wenn auf dieser Ebene personelle Kontinuität herrscht. Gleiches gilt für die Wissenschaft; Kompetenzen schrumpfen und Datensätze veralten. Ohne aktive Promotion verkümmern Motivation und Fähigkeiten.

Wer kümmert sich also um die Pflege der Evidenz-Kultur? Wer wirbt entsprechend gegenüber Entscheidungstragenden in Exekutive und Legislative? Wer überzeugt gesellschaftliche Promotoren? Institutionen wie Leopoldina und Bundesbank wären plausible Stimmen mit der Chance auf Gehör.

Literatur

Allensbach (2009). „Bessere Bedingungen für Familien“. Allensbacher Archiv, IfD-Umfrage 11032.

Bonin, H. u.a. (2014). „Familienpolitische Maßnahmen in Deutschland“. Vierteljahreshefte zur Wirtschaftsforschung 83, 5-11.

Bonin, H. u.a. (2013). „Zentrale Resultate der Gesamtevaluation familienbezogener Leistungen“. DIW Wochenbericht 40, 3-13.

Mackroth, P., Ristau, M. (2002). „Die Rückkehr der Familie“. Berliner Republik 6, 43-47.

Mohn, L., von der Leyen, U. (Hrsg.) (2007). „Familie gewinnt“. Gütersloh.

Prognos (2014). „Gesamtevaluation der ehe- und familienbezogenen Maßnahmen und Leistungen in Deutschland. Endbericht“. Berlin.

Ristau, M. (2005). „Der ökonomische Charme der Familie“. Aus Politik und Zeitgeschichte 23-24, 16-23.

Ristau, M. (2006). „Auf dem Weg zu einer nachhaltigen Familienpolitik“. ifo Schnelldienst 9, 6-9.

Ristau, M. (2015). „Familienpolitik 2000-2013“. Kasseler Diskussionspapiere i3.

Rürup, B. (2009). „Vom Elend der wissenschaftlichen Politikberatung“. Manuskript Last Lecture, Göttingen, 02. Juli 2009.

Schmidt, R., Mohn, L. (Hrsg.) (2004). „Familie bringt Gewinn“. Gütersloh. 2004.

Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.) (2017). „Familienreport 2017“.

<https://www.bmfsfj.de/blob/119524/f51728a14e3c91c3d8ea657bb01bbab0/familienreport-2017-data.pdf> [Zugriff 03.09.2018]

Evidenzbasierung in der Familienpolitik des Bundes

Von PETRA MACKROTH

Korrespondenzadresse:

Bundesministerium für Familie, Senioren, Frauen und Jugend

Glinkastr. 24

10117 Berlin

AL2@bmfsfj.bund.de

Die zentrale Ausgangsfrage lautet: „Wie kann evidenzbasierte Politik gelingen?“ Zur Beantwortung kann und sollte auf die Erfahrungen – und es sei hier schon vorweggenommen – guten Erfahrungen aus der Familienpolitik zurückgegriffen werden.

Die Familienpolitik hat vor ungefähr 15 Jahren, angefangen mit Renate Schmidt und fortgesetzt durch Ursula von der Leyen, über Manuela Schwesig bis heute den Weg verfolgt, nach wissenschaftlich belegbaren Erkenntnissen über die Wirkung von Familienleistungen zu suchen und diese für die weitere Arbeit zu berücksichtigen. Dieser Ansatz führte zur sog. Gesamtevaluation der Ehe- und Familienleistungen, auf die später näher eingegangen wird. Die Idee der wissenschaftsbasierten Familienpolitik ist bis heute wesentliches Merkmal der Weiterentwicklung unserer Familienpolitik. Auch darauf wird im Folgenden näher eingegangen.

Ein erster Gelingensfaktor evidenzbasierter Politikgestaltung wurde schon angedeutet: die Bereitschaft bzw. der Wille der führenden politischen Verantwortungsträger, wissenschaftliche Erkenntnisse als eine Grundlage für politische Entscheidungen ernsthaft nutzen zu wollen. Wenn die politische Führung nicht hinter diesem Ansatz steht, wird evidenzbasierte Politik nur schwerlich gelingen.

Ein weiterer wesentlicher Gelingensfaktor evidenzbasierter Politik liegt in der klaren Zielorientierung auf Seiten der Politik und einer eindeutigen Zielbeschreibung für die wissenschaftliche Untersuchung. Denn ohne eine klare Zielformulierung lässt sich eine Evidenzbasierung und Wirkungsanalyse nicht sinnvoll, nicht zielgerichtet vornehmen. Zudem sollte sich Politik an Zielen ausrichten, nicht entlang von Leistungen, Regelungen und Instrumenten. Sie sind das Mittel zur Zielerreichung.

In der Familienpolitik haben wir mit dem 7. Familienbericht den entscheidenden Schritt 2005 vollzogen: Ziele wurden klar definiert. Darauf aufbauend wurde gefragt, wie diese Ziele bestmöglich – auch im Sinne ökonomischer, fiskalischer Effizienz – erreicht werden können. Die Orientierung auf ein bestimmtes Ziel – im Gegensatz zur Instrumentenfokussierung – weitet den Blick und lässt das Spektrum möglicher Instrumente überhaupt erst in Erscheinung treten.

Ein Beispiel: Das Ziel kann nicht sein, das Kindergeld zu erhöhen, weil das Kindergeld so beliebt ist, sondern das Ziel ist es, die wirtschaftliche Stabilität der Familien zu sichern. Und damit kommen dann ganz andere Leistungen in das Blickfeld mit hinein – zum Beispiel der Kinderzuschlag, der Unterhaltsvorschuss oder das Wohngeld. Mit einer Fokussierung auf das Ziel vermeidet man eine Vorfestlegung auf ein Instrument.

Welchen politischen Nutzen hat nun die Evidenzbasierung? Immerhin: Sie kostet Mühen, Geld und Zeit. Der politische Nutzen besteht darin, dass entlang einer Zielorientierung die evidenzbasierte Politikgestaltung wirksamere, effizientere Instrumente zur Zielerreichung aufzeigen kann. Denn in jedem Politikfeld sind die fiskalischen Mittel begrenzt; wenn durch evidenzbasierte Politikgestaltung wirksamere, effizientere Instrumente eingesetzt werden, ist ein höheres Maß an Zielerreichung möglich. Irrwege, falsche und am Ende teure Weichenstellungen können vermieden werden.

Eine weitere Dimension liegt in der Chance, im Vorhinein „gute“ Politik zu begründen, oder auch durch die nachträgliche Evaluierung, „gute“ Politik nachzuweisen. Das ist in unserer (medien-)öffentlichen Welt nicht zu unterschätzen. Evidenzbasierung kann dazu beitragen, mehr und stärkere – weil wissenschaftlich abgesichert bzw. nachgewiesen – Argumente in die öffentliche und politische Diskussion einzubringen und damit die Durchsetzung politischer Initiativen zu erleichtern.

Eine klare Zieldefinition ist also zwingende Voraussetzung für eine Evaluation von politischen Maßnahmen. Das haben wir in der Gesamtevaluation der Ehe- und Familienleistungen ganz klar erfahren.

Denn – und jetzt komme ich zum Miteinander von Politik und Wissenschaft – Wissenschaft muss auch in der Evaluierung wertfrei erfolgen. Sie darf nicht normativ betrieben werden, auch und gerade weil die Politik von Werten und Wertvorstellungen lebt.

Hier ist es entscheidend, dass beide Seiten, Politik und Wissenschaft, ihre Grenzen kennen, gegenseitig anerkennen und im Evaluierungsprozess dauerhaft einhalten. Das ist uns in der Gesamtevaluation sehr gut gelungen.

Dazu haben drei Faktoren wesentlich beigetragen: Zum einen waren an der Gesamtevaluation der Ehe- und Familienleistungen mehr als 70 Wissenschaftlerinnen und Wissenschaftler beteiligt. Diese Pluralität trug zur Neutralität der Ergebnisse bei. Zudem hatten wir für die Evaluation eine sog. Programmgruppe berufen, in der die wissenschaftlichen Erkenntnisse begleitend von Wissenschaftlerinnen und Wissenschaftlern kommentiert wurden, die nicht an dem gerade durchgeführten Forschungsmodul beteiligt waren. Wir hatten damit auch ein Qualitätssicherungsinstrument für die Evaluation. Schließlich war der Prozess von gegenseitiger Offenheit geprägt, Fragen von Wertfreiheit und Normativität sind in den verschiedenen Forschungsmodulen von beiden Seiten angesprochen und Grenzen respektiert worden.

Durch die Gesamtevaluation hat aber nicht nur die Politik an Wissen gewonnen, auch die beteiligten Wissenschaftlerinnen und Wissenschaftler haben neue Erkenntnisse erhalten und neue Erfahrungen gemacht. Dazu hat der interdisziplinäre Ansatz, den wir für die Evaluation bewusst gewählt haben, beigetragen. Hier sind vor allem die demoskopischen Akzeptanzanalysen zu nennen. In diesen Analysen sind zum einen Kenntnis und Bekanntheit von Leistungen abgefragt worden. Denn klar ist, dass Leistungen nur Wirkung entfalten können, wenn sie bekannt sind und genutzt werden. Darüber hinaus haben die Akzeptanzanalysen aber auch Einstellungen der Familien und der Gesamtbevölkerung zu einzelnen Leistungen, zur Erwerbstätigkeit und Arbeitsteilung in den Familien sowie zu grundlegenden Wünschen und Vorstellungen zum (Familien-)Leben erhoben. Diese Erkenntnisse haben geholfen, die theoretischen und empirischen Ergebnisse aller Evaluationsstudien einzuordnen und zu hinterfragen sowie ggf. Nachfragen zu stellen. Nachdem die Akzeptanzanalysen durch die anderen Disziplinen zunächst mit einer gewissen Distanz betrachtet wurden, haben sie sich während des Prozesses der Erkenntnisgewinnung als hilfreich erwiesen. Für die Politik, die auf die Wahrnehmung ihrer Leistungen durch die Bevölkerung in besonderem Maße angewiesen ist, hatten die Akzeptanzanalysen einen zusätzlichen Wert. Denn Politik kann dauerhaft nicht funktionieren,

wenn sie sich über in der Bevölkerung breit getragene Einstellungen und Werthaltungen hinwegsetzt.

Zum Gelingen einer evidenzbasierten Politik braucht es also zusammenfassend:

- die Bereitschaft der politischen Führung zur wissensbasierten Politikgestaltung
- einen klar erwarteten Nutzen im Hinblick auf ein höheres Maß an Zielerreichung bzw. eine erfolgreichere Politikgestaltung
- die Definition politischer Ziele
- das gegenseitige Anerkennen der unterschiedlichen Arbeitsansätze: Normativität in der Politik und Wertefreiheit in der Wissenschaft.

Es ist mittlerweile gute und geübte Praxis, dass in der Familienpolitik im Rahmen der Gesetzgebung Evaluationen von Leistungen gleich festgelegt werden. Ich bin überzeugt, dass uns dieser Ansatz zu einer besseren Familienpolitik geführt hat und weiterhin führen wird.

Ich nenne hier nur ein Beispiel aus der jüngeren Vergangenheit:

Wir haben das Elterngeld entlang der Erkenntnisse weiterentwickelt, die wir aus Statistiken und Befragungen gezogen und die wir bei der Überprüfung des Ziels, durch diese Leistung die wirtschaftliche Stabilität der Familien in der Familiengründungsphase zu erhalten und zugleich die Vereinbarkeit für Mütter und Väter zu verbessern, gewonnen haben.

Wir haben gesehen, dass viele Elternteile, Mütter wie Väter, vor Ablauf des Elterngeldbezugs in Teilzeit in den Beruf zurückkehren. Das haben wir unterstützt mit der Möglichkeit, bei Teilzeit länger im Elterngeldbezug und damit unterstützt zu sein. Heute haben wir eine Mitwirkungsdauer von acht Monaten durch die Väter im sog. ElterngeldPlus gegenüber nur drei Monaten bei Vollausstieg, denn Väter wollen sich auch um ihre Kinder kümmern. Und wir sehen einen früheren Wiedereinstieg der Mütter in den Beruf, weil sie von ihren Partnern dabei unterstützt werden und die Betreuung der Kinder arbeitsteilig gelingt. Das sichert die gesamte Familie nachhaltig.

Die Evaluationen dienen nicht nur dazu, die Wirksamkeit und Effizienz der Instrumente zu überprüfen, sie helfen auch, unsere Ziele zu überprüfen und weiterzuentwickeln. Denn Gesellschaft entwickelt sich weiter. Beispielsweise war und ist uns die gute Vereinbarkeit von Familie und Beruf ein wichtiges Ziel. Verschiedene Leistungen sorgen dafür. Doch die Wissenschaft hat uns gezeigt: Vereinbarkeit hat auch eine Geschlechterkomponente. Es geht um gute Chancen für Mütter und Väter in Beruf und Familie. Wissenschaftliche Studien haben Nachteile der Frauen in der Arbeitsbiographie nach der Familienphase aufgezeigt. Sie haben auch auf die bessere Bindung von Vätern und Kindern hingewiesen, wenn die Väter früh an der Kinderbetreuung beteiligt waren – was diese im Übrigen auch wollen.

Seit der letzten Legislaturperiode haben wir die partnerschaftliche Vereinbarkeit als weitergehendes Ziel der Vereinbarkeit definiert. Im Interesse der Frauen, Männer und Kinder. Dieser Trend in den jungen Familien ist unübersehbar – er wird prägend sein. Vereinbarkeitsbedingungen müssen zukünftig deshalb den Gesichtspunkt der Partnerschaftlichkeit aufgreifen.

Die Evaluationen der Familienpolitik belegen, dass wir in den vergangenen 10 – 15 Jahren große Fortschritte gemacht haben. Unsere evidenzbasierte Politik ist erfolgreich.

Zehn Thesen für eine erfolgreiche evidenzbasierte Politikberatung: Das Beispiel der Familienpolitik

Von C. KATHARINA SPIESS

Korrespondenzadresse:

Deutsches Institut für Wirtschaftsforschung (DIW) Berlin

Mohrenstr. 58

10117 Berlin

kspiess@diw.de

Viele Jahre hatte eine evidenzbasierte Familienpolitik und eine empirisch ausgerichtete Familienforschung keine bzw. eine nur geringe Bedeutung. Mit Anfang dieses Jahrhunderts nahm jedoch die empirische Evaluation von familienpolitischen Maßnahmen immer weiter zu. Dies geht mit einem Bedeutungszuwachs der empirisch fundierten Wirtschaftsforschung und damit auch der empirisch fundierten Familienökonomie einher. Mit immer mehr und besseren Mikrodaten konnten und können familienpolitische Maßnahmen für Gesamtdeutschland evaluiert werden. Auch die international vorangetriebene Weiterentwicklung von Methoden der Evaluationsforschung hat signifikant dazu beigetragen, dass die Bedeutung der evidenzbasierten Familienpolitik zugenommen hat.

Auf der Angebotsseite einer evidenzbasierten Politikberatung, also auf der Seite der Forschung, hat insbesondere auch die empirische Arbeitsmarktforschung zu einem erheblichen Schub in der empirischen Evaluationsforschung beigetragen. Dies hat sich auch auf die empirisch arbeitende Familienökonomie ausgewirkt, welche basierend auf repräsentativen Mikrodaten vielfach die Methoden der Arbeitsmarktevaluation auf die Familienpolitik übertragen hat. Etwas versetzt ist auch auf der Nachfrageseite, also der Politik und Administration im Bereich der Familienpolitik, ein zunehmendes Interesse an einer Evidenzbasierung der eigenen Politik hinzugekommen. Evidenzbasierung wurde als ein zusätzliches Argument zur Begründung von familienpolitischen Reformen nachgefragt. Aber auch im Vorfeld von Reformen bzw. der Einführung neuer familienpolitischer Maßnahmen wurde eine Evidenzbasierung als zunehmend notwendig erachtet, um die Wirkung alternativer

Maßnahmen abzuschätzen und damit Politik zu gestalten. Ein Beispiel hierfür ist die Forschung zu den Wirkungen unterschiedlicher Alternativen eines Elterngeldes.¹ Unabhängig davon konnten mit einer evidenzbasierten Familienpolitik andere Mehrheiten für familienpolitische Reformen gewonnen werden als eine Familienpolitik, die sich weitgehend auf normativ geprägte Diskurse stützt. In diesem Kontext können Analysen zu den Notwendigkeiten und Wirkungen des Ausbaus einer Kindertagesbetreuung genannt werden, welche beispielsweise ökonomische Gründe für den Ausbau der Kindertagesbetreuung lieferten.² Die Verabschiedung des Kinderförderungsgesetzes im Jahr 2008 war somit auch das Ergebnis einer politischen Kommunikation, die sich einer Vielzahl von Argumenten bediente. Außerdem wurde mit diesem Gesetz erstmalig bei einem Gesetz mit familienpolitischen Bezug eine regelmäßige Evaluation der beschlossenen Maßnahmen festgeschrieben.³

Allerdings kann insgesamt sowohl auf der Anbieterseite und insbesondere auf der Nachfrageseite kein konstantes Interesse an einer Evidenzbasierung der Familienpolitik festgestellt werden – vielmehr sind auch hier konjunkturelle Schwankungen zu beobachten. Vor diesem Hintergrund soll es im Folgenden darum gehen, zehn Thesen zu einer erfolgreichen evidenzbasierten Familienpolitik zu diskutieren. Dies geschieht primär auf der Basis meiner Erfahrungen im Bereich der evidenzbasierten familienpolitischen Beratung seit Anfang dieses Jahrhunderts. Insbesondere basieren die Thesen auf Erfahrungen im Rahmen der Gesamtevaluation ehe- und familienbezogener Leistungen, einem großen Projekt der Jahre 2009 bis 2014, das ich in unterschiedlichen Phasen und Funktionen als Familienforscherin begleiten konnte.⁴

Unabhängig davon handelt es sich bei der Gesamtevaluation ehe- und familienbezogener Leistungen in mehrfacher Hinsicht um ein Lehrstück der evidenzbasierten Politikberatung. Die Gesamtevaluati-

1 Vgl. z.B. BÜCHNER ET AL. 2006; SPIEB, WROHLICH 2006.

2 Vgl. z.B. SPIEB ET AL. 2002 und SPIEB, WROHLICH 2005.

3 In den Jahren 2010 bis 2015 wurden Daten zur Kindertagesbetreuung jährlich in einem Bericht zum Kinderförderungsgesetz („KiföG-Bericht“) veröffentlicht. Seit 2016 werden die jeweils letztverfügbaren Daten zum bundesweiten Ausbaustand und Bedarf in der Kindertagesbetreuung in der Publikation „Kindertagesbetreuung Kompakt“ veröffentlicht.

4 Für den Abschlussbericht der Gesamtevaluation ehe- und familienbezogener Leistungen: BÖHMER ET AL. 2014.

on untersuchte im Auftrag des Bundesministeriums der Finanzen und des Bundesministeriums für Familie, Senioren, Frauen und Jugend ab Herbst 2009 die Wirkung zentraler ehe- und familienbezogener Leistungen. Dabei ist besonders hervorzuheben, dass dieses große Vorhaben von zwei Ressorts mitverantwortet wurde. Insgesamt wurden familienbezogene Leistungen, für die im Jahr 2010 153 Mrd. Euro ausgegeben wurden, evaluiert. Die Evaluation umfasste 12 Forschungsmodule, an denen über 70 Wissenschaftlerinnen und Wissenschaftler aus unterschiedlichen Disziplinen, mit einem Schwerpunkt von Ökonominen und Ökonomen, mitwirkten. Diese Evaluation war in ihrem Charakter national als auch international einmalig - insofern erweckte sie auch bei internationalen Akteuren, wie beispielsweise der OECD in Paris, großes Interesse. Was lässt sich thesenhaft vor diesem Hintergrund für eine erfolgreiche Evaluation (familien-)politischer Maßnahmen bzw. einer evidenzbasierten Politikberatung festhalten?

1. These: Eine Evaluation setzt Klarheit darüber voraus, was die Ziele der Politik bzw. Maßnahmen sind, welche evaluiert werden sollen (Zielorientierung).

Bei der Gesamtevaluation ehe- und familienbezogener Leistungen wurde sehr deutlich, wie wichtig es ist, sich im Vorfeld auf die Ziele zu einigen, die für das Politikfeld relevant sind. Im Kontext der Familienpolitik war dies insofern neu, da bisher kein expliziter Zielkatalog der deutschen Familienpolitik vorlag. Vielmehr wurden mit unterschiedlichen Maßnahmen unterschiedliche Ziele verfolgt. Die Ziele, welche für die Gesamtevaluation relevant waren, gingen auf das Konzept einer „Nachhaltigen Familienpolitik“ zurück, wie sie 2005 erstmalig von Malte Ristau benannt wurde.⁵

Eine entsprechende Zielbenennung für eine Evaluation muss klar und eindeutig sein und sollte von der Politik und nicht der Wissenschaft vorgenommen werden. Sie ist für die Politik allerdings mit Risiken verbunden, da der Erfolg ihrer Politik an der Erreichung ihrer Ziele festgemacht werden kann. Eine klare und eindeutige Zielfestlegung ermöglicht darüber hinaus die Identifikation von Zielkonflikten, was ein

5 RISTAU 2018.

weiteres Risiko für die Politik sein kann. Allerdings ist auch zu bedenken, dass Zielkonflikte eines historisch gewachsenen Sets von Maßnahmen nahezu unumgänglich sind. In Abhängigkeit politischer Mehrheiten werden unterschiedliche Maßnahmenbündel mit unterschiedlichen Zielsetzungen implementiert, welche bei veränderten Mehrheiten i.d.R. nicht rückgängig gemacht werden.

Des Weiteren erfordert eine Zielbenennung für eine Evaluation die Identifikation von Hauptzielen bzw. Oberzielen, da nur ein begrenzter Zielkatalog Gegenstand von Evaluierungen sein kann. Eine Zielhierarchisierung ist Aufgabe der Politik, welche Evaluationen in Auftrag gibt. Von zentraler Bedeutung ist, dass der vor Beginn der Evaluation festgesetzte Zielkatalog nicht während der Evaluation verändert werden darf. Im Rahmen der Gesamtevaluation ehe- und familienbezogener Leistungen wurden vorab die folgenden vier Ziele als Hauptziele identifiziert: Erstens die Sicherung der wirtschaftlichen Stabilität von Familien, zweites eine verbesserte Vereinbarkeit von Familie und Beruf, drittens die Förderung und das Wohlergehen von Kindern sowie viertens die Erfüllung von Kinderwünschen. Während des Evaluationsprozesses wurde unter der Bundesfamilienministerin Kristina Schröder allerdings der Versuch gemacht, das Ziel der „Wahlfreiheit“ als ein weiteres familienpolitisches Ziel der Evaluation festzulegen. Eine solche Intervention während des Evaluationsprozess war insbesondere für die Außenwirkung des Gesamtprozesses sehr schädlich. Dadurch ist der Eindruck des Versuchs einer politischen Einflussnahme entstanden.

2. These: Die Evaluation von Maßnahmen setzt eine Verständigung auf die Operationalisierung von Zielen voraus (Zieloperationalisierung).

Ein weiterer zentraler Bestandteil einer erfolgreichen Evaluation ist eine Verständigung auf die Operationalisierung von Zielen. Auch dieser Prozess sollte idealerweise der eigentlichen Evaluation vorgeschaltet sein und transparent gemacht werden. Eine Operationalisierung und Quantifizierung der Ziele ist insbesondere im Bereich der Familienpolitik keinesfalls trivial, wenn zum Beispiel an das Ziel der Förderung von Kindern oder auch die wirtschaftliche Stabilität der Familie gedacht wird. Denn u.a. ist eine Zielerreichung auch davon abhängig, in welchem Zeitraum – kurz-, mittel- oder langfristig – sie erfolgen kann.

Des Weiteren sollten sich Politik und Wissenschaft im Vorfeld auf eine Operationalisierung von Zielen verständigen, damit diese nicht unter dem Verdacht steht, erst im Nachhinein ergebnisabhängig definiert zu werden. Entsprechende Diskussionen müssen deshalb frühzeitig stattfinden und es sollte transparent gemacht werden, was deren Ergebnis ist. Eine entsprechende Operationalisierung kann auch im Kontext von Machbarkeitsstudien, Vorstudien oder auch auf der Basis von Pretests stattfinden.

3. These: Das Evaluationsziel sollte die Überprüfung der Effektivität von Maßnahmen sein und zwar im Sinne einer Identifikation der kausalen Wirkung derselben (Kausalität).

Auch wenn es der ökonomischen Evaluationsforschung inzwischen immanent ist, dass es dabei um eine Identifikation kausaler Effekte geht, so ist es wichtig, dies auch für eine evidenzbasierte Politikberatung festzuhalten. Differenzierte deskriptive Analysen von Zusammenhängen können einen Erkenntnisgewinn haben, sind aber häufig nur der erste Schritt. Letztlich sind mikroökometrische Verfahren zur Identifikation kausaler Effekte notwendig. Für die Wissenschaft besteht die Herausforderung darin, dies der Politik zu kommunizieren und die Besonderheiten spezifischer Verfahren hervorzuheben, welche für die Interpretation und insbesondere Verallgemeinerung der Ergebnisse von Bedeutung sind. So müssen insbesondere quasi-experimentelle Studien das Problem der externen Validität ansprechen, während Studien auf der Basis struktureller Modelle insbesondere die zugrundeliegenden Annahmen in Hinblick auf politische Schlussfolgerungen adressieren sollten. Der Politik wiederum muss klar sein, dass sozialwissenschaftliche Evaluationsansätze grundsätzlich keine naturgesetzlichen Aussagen machen können und Ergebnisse stets auf der Basis bestimmter Annahmen in Zeit, Raum und Kontext zu interpretieren sind.

Vor diesem Hintergrund gilt es, die Effektivität politischer Maßnahmen idealerweise für die Mikro-, Meso- und Makroebene zu identifizieren. Es müssen die Effekte auf unterschiedliche Gebietskörperschaften sowie kurz-, mittel- und langfristige Wirkungen betrachtet werden. Inzwischen ist die Analyse von Effektheterogenitäten ein wichtiger zentraler Bestandteil entsprechender Wirkungsstudien.

4. These: Ein weiteres Ziel sollte – sofern möglich – die Überprüfung der Effizienz von Maßnahmen sein.

Aus einer ökonomischen Perspektive ist neben der Erfassung der Effektivität insbesondere auch die Frage der Effizienz von Bedeutung. Dies setzt allerdings eine differenzierte Erfassung der tangiblen und intangiblen Kostenströme voraus, die mit der zu bewertenden Maßnahme verbunden sind. Eine vollständige Kostenerfassung ist ressourcenintensiv. In der Erfassung der Nutzenströme liegt eine weitere Herausforderung. Wenn von einer monetären Bewertung der Nutzenströme abgesehen werden soll, kann eine relative Effizienzerfassung im Rahmen einer Kosten-Effektivitätsanalyse sinnvoll sein. Eine Kosten-Nutzen-Analyse im engeren Sinne setzt allerdings eine Monetarisierung von Nutzenströmen voraus, was insbesondere für die Familienpolitik sehr anspruchsvoll ist. Dies wird deutlich, wenn an eine Monetarisierung von Entwicklungsfortschritten von Kindern oder auch der Erfüllung von Kinderwünschen gedacht wird. Hinzu kommt, dass auch hier langfristige Betrachtungen notwendig sind und idealerweise Nutzenströme über den gesamten Lebenszyklus abzuschätzen sind. In diesem Fall müssen Diskontierungsraten festgelegt werden, mit denen künftige Nutzen- und auch Kostenströme auf den Betrachtungszeitraum abdiskontiert werden.

5. These: Eine Evaluation sollte auch die Seite der Betroffenen betrachten und zwar in Hinblick auf Bekanntheitsgrad, Detailkenntnisse und Akzeptanz von Maßnahmen (Wissen von potentiell Betroffenen).

Eine Evaluation sollte auch die Sicht der von der Maßnahme betroffenen Individuen bzw. Akteursgruppen einbeziehen. Dies betrifft insbesondere das vorhandene Wissen der Betroffenen in Hinblick auf die Existenz von Maßnahmen, Zugangsvoraussetzungen, Antragsverfahren, der Angebotsvielfalt etc. Für eine Evaluation ist somit auch relevant, inwiefern diesbezügliche Informationsasymmetrien vorliegen. Auch im Rahmen der Gesamtevaluation ehe- und familienbezogener Leistungen wurden sogenannte Akzeptanzanalysen durchgeführt, welche auf der Basis demoskopischer Erhebungen Wissen über familienpolitische Maßnahmen erfasst haben.

6. These: Die Übertragbarkeit der Evaluationsergebnisse sollte thematisiert werden (externe Validität).

Für die Interpretation der Ergebnisse und die politische Verwertbarkeit der Evaluationsergebnisse ist es fernerhin von Bedeutung, die externe Validität der Evaluationsergebnisse zu thematisieren. Dabei geht es nicht nur um die bereits oben diskutierte externe Validität in Hinblick auf bestimmte Methoden, sondern auch um die externe Validität z.B. in Hinblick auf die verwandten Daten. Letztlich sollte sich eine evidenzbasierte Politik auf eine Vielzahl von Studienergebnissen stützen. Es sollten nur jene Politikempfehlungen aus der Wissenschaft umgesetzt werden, welche sich unabhängig von Daten und Methoden replizieren lassen. Es bedarf also der Redlichkeit der Wissenschaft, auf eventuelle Grenzen der Übertragbarkeit hinzuweisen und es bedarf der Redlichkeit der Politik, dies ebenfalls in der politischen Kommunikation zu berücksichtigen. Hilfreich ist dabei eine Evaluationskompetenz der Administration, welche die Politik bei der Interpretation der Ergebnisse berät und welche im Umgang mit vielfältigen Ergebnissen geschult ist. In Hinblick auf die Prüfung der externen Validität ist zum einen die Förderung von wissenschaftlichen Replikationsstudien von Bedeutung. Zum anderen können Vergleiche mit internationalen Studien Hinweise darauf geben, welche Evaluationsergebnisse institutionen- und/oder kontextgebunden sind.

7. These: Die Evaluation muss von unterschiedlichen wissenschaftlichen Institutionen durchgeführt werden (Neutralität und Wettbewerb).

Evaluationen sollten von der Politik öffentlich ausgeschrieben sein und die Wissenschaft sollte sich in wettbewerblichen Verfahren dafür qualifizieren. In diesem Wettbewerb muss es um das effizienteste Evaluationsdesign, die besten Daten, Methoden und disziplinären Zugänge gehen. Ein solcher Wettbewerb sollte allerdings moderiert sein, das heißt, er kann sich nicht selbst überlassen werden. Im Rahmen der Gesamtevaluation ehe- und familienbezogener Leistungen hat diese Aufgabe ein privater Akteur übernommen, der sich in einem wettbewerblichen Verfahren dafür qualifiziert hat.

In diesem Kontext sind auch Fragen der Governance von Evaluationsprozessen differenziert zu betrachten und im Vorfeld zu überlegen.

Von zentraler Bedeutung ist, dass die Akteure, welche die Evaluation durchführen, neutral sind und keinen politischen Handlungslogiken unterliegen. Bei groß angelegten Evaluationen, wie der Gesamtevaluation, sollten mehrere Forschungseinrichtungen beteiligt werden.

8. These: Erfolgreiche Evaluationen benötigen erhebliche Ressourcen (Zeit und Geld).

Eine effektive Evaluation setzt erhebliche finanzielle und zeitliche Ressourcen voraus. Bei der Frage, wer diese aufzubringen hat, ist zum einen an eine öffentliche Finanzierung durch die Politik bzw. Administration zu denken. Sofern mehrere Ressorts oder politische Akteure an der Finanzierung beteiligt sind, sollten sie sich über die gemeinsamen Ziele im Vorfeld verständigen. Zum anderen kann eine evidenzbasierte Politikberatung bzw. eine Wirkungsforschung auch von der Wissenschaft selbst finanziert werden. Beide Finanzierungsmöglichkeiten greifen letztlich auf Gelder öffentlicher Haushalte zurück. Wenn eine Finanzierung durch die Wissenschaft infrage käme, könnte z.B. an Forschungseinrichtungen gedacht werden, welche sich als Einrichtungen der Leibniz-Gemeinschaft dem Grundsatz „*theoria cum praxi*“ verpflichten. Eine über den Grundhaushalt dieser Forschungseinrichtungen finanzierte Evaluationsforschung darf ebenso wie eine von Ministerien oder anderen öffentlichen Auftraggebern finanzierte Forschung nicht zu Lasten von Wissenschaftsstandards gehen. Dabei ist beispielsweise auch eine transparente Darstellung *aller* Ergebnisse zu nennen.

Neben ausreichend finanziellen Ressourcen sind ausreichend zeitliche Ressourcen notwendig. Auch ein hoher Einsatz zeitlicher Ressourcen kann mit Zielkonflikten verbunden sein, die es explizit mit zu bedenken gilt. Für die Wissenschaft bzw. die Angebotsseite einer empirischen Wirkungsforschung muss bedacht werden, dass die Wissenschaft sich in der Bewertung von Individuen und Einrichtungen an Forschungsrankings orientiert. Für diese Rankings sind z.B. Publikationen in hochwertigen Fachjournals von größter Bedeutung. Transferpublikationen haben einen sehr viel geringeren Stellenwert. Für die Nachfrageseite ist es häufig wichtig, zeitnah Ergebnisse vorliegen zu haben, da eine (geplante) Maßnahme beispielsweise gegenüber der Öffentlichkeit oder der Opposition gerechtfertigt werden muss. Hinzu kommt, dass die Politik sich an Legislaturperioden orientiert, welche in der Handlungslogik der

Wissenschaft keine Bedeutung haben. Diese unterschiedlichen Logiken sollten beiden Seiten bewusst sein, um das gegenseitige Verständnis zu prägen.

9. These: Ein öffentlicher Zugang zu allen Evaluationsergebnissen bzw. eine breite Diskussion derselben ist essentiell für eine erfolgreiche Evaluation (transparente Ergebnisdarstellung).

Eine erfolgreiche Evaluation, welche mit öffentlichen Geldern finanziert wird, sollte in Hinblick auf ihre Ergebnisse transparent gemacht werden und in der Fachöffentlichkeit hinsichtlich ihrer Implikationen diskutiert werden. Vielfach bedeutet dies für die Wissenschaft eine Übersetzung der Ergebnisse in eine Transfersprache, die es ermöglicht, die Ergebnisse verständlich und in ihrem Detail darzulegen. Auf Seiten der Politik und Administration ist hier ein Know-how im Bereich der wissenschaftlichen Evaluationsforschung hilfreich, während auf Seiten der Wissenschaft ein Verständnis von Prozessen der politischen Kommunikation hilfreich ist. Ein entsprechendes gegenseitiges Verständnis kann in Workshops und Fachgesprächen vertieft werden oder Folge eines expliziten Personaltransfers oder Personalaustausches in beide Richtungen sein. So könnten zum Beispiel Wissenschaftlerinnen und Wissenschaftler ein Sabbatical auch für entsprechende Aufenthalte bei der Administration nutzen und die Administration könnte Sabbaticals für einen befristeten Aufenthalt in Forschungseinrichtungen nutzen. Eine transparente Ergebnisdarstellung setzt auch voraus, dass alle Evaluationsergebnisse offengelegt werden. Geschieht dies nicht, wird riskiert, dass die veröffentlichten Ergebnisse keine Akzeptanz finden. Letzteres deshalb, da der Eindruck einer Beeinflussung der Ergebnisdarstellung entstehen kann.

10. These: Die Evaluationsergebnisse sollten sich in politischen Maßnahmen bzw. Reformen widerspiegeln (Wirksamkeit).

Die Wirksamkeit der Evaluationsforschung ist ein weiterer zentraler Bestandteil einer erfolgreichen Evaluation. Sofern die Ergebnisse einen Handlungsbedarf identifizieren, sollte dieser angegangen werden. Allerdings liegen auch hier unterschiedliche Herausforderungen vor: Insbesondere bei Evaluationen, welche einen längeren Zeitraum umfas-

sen, kann es sein, dass mit einem Wechsel in den politischen Mehrheiten Veränderungen in der politischen Prioritätensetzung einhergegangen sind und z.B. die evaluierten Politikbereiche nicht mehr als relevant erachtet werden. Des Weiteren liegt eine Herausforderung darin, dass insbesondere auf Seiten der Ökonomie keine Kompetenz im Bereich der Implementierung neuer Maßnahmen vorliegt. Hinzu kommt, dass die Wissenschaft keine eigenen Ressourcen hat, um Implementationsprozesse zu begleiten und eine entsprechend wissenschaftlich begleitete Implementation weitere finanzielle und zeitliche Ressourcen mit sich bringen würde. Letzteres wird häufig bei der Abschätzung relevanter Ressourcen im Vorfeld nicht bedacht. Ohne eine erfolgreiche Implementationsstrategie von Reformmaßnahmen sind Evaluationen allerdings nur bedingt erfolgreich. Somit zeigt sich, wie vielfältig und aufwendig entsprechende Evaluationen sein können. Eine evidenzbasierte Familienpolitik ist jedoch ebenso wie eine evidenzbasierte Arbeitsmarkt-, Gesundheits- oder Wirtschaftspolitik von elementarem Interesse, wenn Politik zugunsten der Bevölkerung gemacht werden soll.

Literatur

Büchner, C., Haan, P., Schmitt C., Spieß, C.K., Wrohlich, K. (2006). „Wirkungsstudie Elterngeld“. DIW Berlin Politikberatung kompakt 18.

Böhmer, M., Ehrentraut, O., Heimer, A., Henkel, M., Ohlmeier, N., Poschmann, K., Schmutz, S., Weisser, J. (2014). „Gesamtevaluation der ehe- und familienbezogenen Maßnahmen und Leistungen in Deutschland. Endbericht“. Basel. <https://www.bmfsfj.de/blob/93954/25490622c47497e47acbcfa797748cfb/gesamtevaluation-der-ehe-und-familienbezogenen-massnahmen-und-leistungen-data.pdf> [Zugriff 7. Juli 2018]

Ristau, M. (2005). „Der ökonomische Charme der Familie“. Aus Politik und Zeitgeschichte 23-24, 16-23.

Spieß, C.K., Schupp, J., Grabka, M., Haisken-De New, J.P., Jakobeit, H., Wagner, G.G. (2002). „Abschätzung der Brutto-Einnahmeneffekte öffentlicher Haushalte und der Sozialversicherungsträger bei einem Ausbau von Kindertageseinrichtungen“. Baden-Baden: Nomos Verlag.

Spieß, C.K., Wrohlich, K. (2005). „Wie viele Kinderbetreuungsplätze fehlen in Deutschland?“. DIW-Wochenbericht 14, 223-227.

Spieß, C.K., Wrohlich, K. (2006). „Elterngeld: Kürzere Erwerbspausen junger Mütter erwartet“. DIW-Wochenbericht 48, 689-694.

Die Gesamtevaluation ehe- und familienbezogener Leistungen: Meilenstein auf dem Weg zu einer evidenzbasierten Familienpolitik in Deutschland

Von HOLGER BONIN

Korrespondenzadresse:

Forschungsinstitut zur Zukunft der Arbeit GmbH (IZA)

Schaumburg-Lippe-Str. 5-9

53113 Bonn

bonin@iza.org

Nach der groß angelegten Evaluation der aktiven Arbeitsmarktpolitik im Umfeld der Hartz-Reformen war die Familienpolitik das zweite Maßnahmenfeld im Zuständigkeitsbereich des Bundes, auf dem die vorhandene Vielzahl von Leistungen einer umfassenden Bewertung unterzogen wurde. Einen wichtigen Anstoß dazu gab der 7. Familienbericht, der einen Perspektivwechsel von Familienpolitik als sozialer Ausgleichspolitik hin zu Familienpolitik als nachhaltiger Unterstützung von Eltern und Kindern über den Lebenslauf unterstützte. Die Sachverständigenkommission legte in ihrem Bericht offen, dass eine auf Nachhaltigkeit ausgerichtete Familienpolitik nicht ohne Evaluation der eigenen Maßnahmen auskommen kann.¹

In der Folge beauftragte das Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) gemeinsam mit dem Bundesministerium der Finanzen (BMF) im Herbst 2009 die „Gesamtevaluation der ehe- und familienbezogenen Leistungen in Deutschland“. Dieses über vier Jahre laufende Großvorhaben hat durch eine vergleichende Bewertung von Effektivität und Effizienz der zentralen Instrumente der Familienförderung und ihrer Interaktionen eine Fülle an Evidenz für eine sowohl zielorientierte als auch wirtschaftliche Politikgestaltung geliefert.² Die Gesamtevaluation ist zweifellos der Meilenstein auf dem Weg zur heute zumindest auf der Ebene des Bundes ziemlich gängig gewordenen Praxis, bei Entscheidungen über die Familienpolitik auch die Wissenschaft und ihre Erkenntnisse mit einzubeziehen.

1 DEUTSCHER BUNDESTAG 2006.

2 PROGNOSE 2014.

Der Prozess der umfassend angelegten Bewertung der Familienförderung wirkt immer noch nach, weil damit die Voraussetzungen für eine stärkere Evidenzbasierung der Familienpolitik in einer Reihe von Dimensionen erheblich gestärkt wurden.

Erstens: Die Familienpolitik hat sich im Vorfeld der Gesamtevaluation auf einen Katalog von Zielgrößen festgelegt, die sie mit ihren Instrumenten verbessern möchte: die wirtschaftliche Stabilität und soziale Teilhabe von Familien, die Vereinbarkeit von Familie und Beruf, die Förderung und das Wohlergehen von Kindern, die Realisierung von Kinderwünschen sowie Nachteilsausgleich zwischen den Familien. Konkretisierungen dieser Art sind für die Politik durchaus schwierig und deshalb keineswegs selbstverständlich. Die vorgenommenen Setzungen haben normativen Charakter und lassen somit hinsichtlich Inhalt und Auswahl je nach Standpunkt Raum für Kritik. So wurden im Zusammenhang mit der Gesamtevaluation eine sich aus dem Vereinbarkeitsziel ergebende Fokussierung auf die Müttererwerbstätigkeit als Maßstab für die Qualität familienpolitischer Maßnahmen sowie das Fehlen des Ziels der Wahlfreiheit von einigen Interessenvertretern zum Teil heftig kritisiert.³ Zudem lässt sich eine Politik leichter als falsch kritisieren, sobald diese mit klaren – und so überprüfbaren – Zielvorgaben versehen wird.

Andererseits benötigt eine optimale Kooperation von Politik und Wissenschaft zum Zweck der Programmevaluation entsprechende Setzungen. Empirische Wirkungsforschung als positive wissenschaftliche Analyse kann diese nicht vornehmen. Ihr obliegt lediglich die Aufgabe, die von der Politik vorgegebenen normativen Ziele so inhaltlich angemessen und transparent wie möglich in messbare Zielvariablen zu überführen.

Zweitens: Als ein Bestandteil der Gesamtevaluation wurde die Datengrundlage für Analysen zur Entwicklung der lebenslauforientierten familienpolitischen Zielgrößen gezielt verbessert. Ein hoher Aufwand zur Verbesserung der Dateninfrastruktur war erforderlich, da im Bereich der Familienpolitik die Inanspruchnahme von Leistungen in Deutschland nicht systematisch administrativ registriert wird. Außerdem umfassen die familienpolitischen Ziele Größen, vor allem die Förderung und das Wohlergehen von Kindern betreffend, die in amtlichen Daten

3 BORCHERT 2013.

allgemein nicht oder nur schlecht abgebildet sind. Die für die Zwecke der Gesamtevaluation angestoßenen Längsschnittbefragungen wurden unter dem Namen „Familien in Deutschland“ (FiD) durchgeführt, und zwar als Ergänzungsstudie zum Sozio-oekonomischen Panel (SOEP), die Zusatzstichproben unter anderem von Großfamilien, Alleinerziehenden und Haushalten mit Kleinkindern umfasst.⁴

Bedeutsam ist, dass es gelungen ist, diese Investition in die Dateninfrastruktur über das Ende der Gesamtevaluation hinaus fortzuführen. Zunächst finanzierte das BMFSFJ eine weitere Erhebungswelle und nach Auslaufen dieser Finanzierung konnten die Haushalte aus FiD in die SOEP-Kernstudie integriert werden. So hat sich die Datenbasis für eine evidenzbasierte Familienpolitik in Deutschland nachhaltig verbreitert. Statistische Unsicherheiten im Hinblick auf familienpolitisch relevante Zielgrößen sind messbar kleiner geworden, wovon Politik wie Wissenschaft profitieren. Da die neuen Daten für Wissenschaftlerinnen und Wissenschaftler sehr leicht zugänglich sind, tragen sie auch jenseits der Auftragsforschung zu mehr Dynamik in der empirischen Familienforschung bei.

Drittens: Die Gesamtevaluation hat neue Kreise von Forscherinnen und Forschern dazu gebracht, sich mit Fragen der Familienpolitik zu befassen und den Aufbau weiterhin nutzbarer Modellinfrastrukturen bei den beteiligten Forschungsinstituten ermöglicht. Um mögliche Herangehensweisen an die beabsichtigte Gesamtevaluation auszuloten, wurde das Vorhaben zunächst mit einer umfangreichen Machbarkeitsstudie vorbereitet. Den Auftrag für diese Studie erhielt ein Konsortium aus drei Wirtschaftsforschungsinstituten in der Leibniz-Gemeinschaft, die zwar mit Methoden der Evaluationsforschung hochgradig erfahren, jedoch mit Ausnahme des DIW Berlin in der wissenschaftlichen Beratung der Familienpolitik nicht besonders aktiv waren. Die Arbeit an der Machbarkeitsstudie sorgte hier für eine Annäherung an die laufenden familienpolitischen Diskurse, aber auch für den Aufbau von Kapazitäten im Hinblick auf die geplante Gesamtevaluation. Die Durchführung einer gründlichen Vorstudie dürfte dazu beigetragen haben, in den sich anschließenden Vergabeverfahren konkurrierende Angebote von mehreren fachlich kompetenten Bietern zu erhalten.

4 SCHRÖDER ET AL. 2013.

Für die Verbreiterung der wissenschaftlichen Basis ebenfalls bedeutsam war die Zerlegung der Gesamtaufgaben der Evaluation in eine Reihe von sich inhaltlich zu einem gewissen Teil überlappenden Modulen. Diese Organisationsweise ermöglichte nicht nur eine Validierung von Befunden durch Abgleich der von verschiedenen Auftragnehmern mit verschiedenen empirischen Verfahren errechneten Resultate. Sie sorgte auch dafür, dass parallel zueinander komplexe verhaltensfundierte Mikrosimulationsmodelle aufgebaut wurden. Diese Modelle spielen im Kontext der Evaluation von Familienleistungen eine besondere Rolle, weil übliche Kontrollgruppenverfahren wegen der teils so gut wie universellen Inanspruchnahme von Leistungen und der komplexen Interaktionen innerhalb des Steuer- und Transfersystems im Bereich der Familienpolitik nämlich nur sehr begrenzt einsetzbar sind.

Für den Aufbau dieser Modellinfrastruktur muss großer Aufwand betrieben werden. Dies erklärt, warum Forscherinnen und Forscher aus dem universitären Umfeld, zumindest bei den Modulen der quantitativen Wirkungsforschung, im Rahmen der Gesamtevaluation kaum zum Zuge kamen. Modellaufbau und Modellpflege lassen sich mit den üblichen Kapazitäten eines Lehrstuhls kaum leisten. Wie die zuvor skizzierte Investition in die Dateninfrastruktur wirkt auch die Investition in die Modellinfrastruktur über den Abschluss der Gesamtevaluation hinaus nach. Es existieren heute erprobte Modellrahmen, die sich im Bedarfsfall relativ schnell für eine Bewertung neuer Maßnahmen in der Familienpolitik aktivieren lassen.

Viertens: Der lange Prozess einer umfassend angelegten und komplexen Evaluation war auch ein tiefgreifender Lernprozess, der das Verständnis der Politik und der Wissenschaft für die Anforderungen und Arbeitsweisen der jeweils anderen Seite nachhaltig gestärkt hat. Es sind Kommunikations- und Vertrauensbeziehungen gewachsen, die heute einen auch informellen Austausch dazu erleichtern, wie Maßnahmen der Familienpolitik zielführend ausgestaltet werden könnten. Zu diesem Austausch trägt auch das von der Prognos AG im Auftrag des BMFSFJ geführte „Kompetenzbüro Wirksame Familienpolitik“ bei, das an der Schnittstelle zwischen Wissenschaft und Politik platziert ist.

In den an der Gesamtevaluation beteiligten Ressorts ist durch die Befassung mit den vielen in diesem Zusammenhang entstandenen Berichten *by doing* ein solides Verständnis über die im Kontext der Fa-

milienpolitik gebrauchten Methoden der Evaluierung, ihre Voraussetzungen und Limitationen entstanden. Dies hat dazu geführt, dass die Anforderungen, die aus den Ressorts heraus gegenüber der Wissenschaft artikuliert werden, höher geworden sind – was die gestellten Aufgaben für Forscherinnen und Forschern nur reizvoller macht. Gleichzeitig haben diejenigen, die wissenschaftlich beratend tätig werden, durch lange Arbeitsbeziehungen mit den Ressorts meist ein gutes Verständnis der Rahmenbedingungen und Restriktionen entwickelt, unter denen sich die Gestaltung der Familienpolitik vollzieht. Dies sorgt dafür, dass sie sich mit einem gewissen Pragmatismus den machbaren Optionen zuwenden und nicht beim prinzipiell Wünschenswerten verharren.

Zudem sind viele von denen, die Familienpolitik evaluieren und darüber mit der Politik im Dialog stehen, in der Lage, ihre Erkenntnisse und Einschätzungen zu einem gesellschaftlich kontrovers betrachteten Gegenstand, auf den eine Vielzahl von Interessenverbänden, aber auch die Bürgerinnen und Bürger sehr aufmerksam blicken, differenziert und neutral zu kommunizieren. Diese Fähigkeit konnte sich besonders in dem Moment bewähren, als die Gesamtevaluation nach ihrem Abschluss wegen eines von der zu dieser Zeit zuständigen Familienministerin Kristina Schröder initiierten politischen Berichts, der zentrale Befunde entstellend interpretierte, zu scheitern drohte. In dieser Lage konnte auch mit Hilfe einer aktiven und glaubwürdigen medialen Kommunikation der Wissenschaft abgewendet werden, dass Auftragsforschung, die politisch ungewollte Evidenz hervorbringt, „in die Schublade wandert.“ Dieses Beispiel lehrt im Übrigen die wohl elementarste Voraussetzung, damit Politik und Wissenschaft optimal kooperieren können: die bedingungslose Wahrung der guten wissenschaftlichen Praxis und damit der prinzipiellen Öffentlichkeit von Forschung.

Wie es scheint, kommt heute – fünf Jahre nach dem Ende der Gesamtevaluation – die mit ihr verankerte evidenzbasierte Herangehensweise an Familienpolitik vor allem dort zum Tragen, wo es darum geht, das Steuer- und Transfersystem für Eltern und ihre Kinder umzugestalten und die potenziellen Wirkungen von Reformoptionen auf die politischen Ziele im Vorfeld zu bewerten. So ließ die Bündnis 90/Die Grünen-Bundestagsfraktion im Vorfeld der jüngsten Bundestagswahl verschiedene Reformoptionen für eine verbesserte materielle Absicherung von Familien mit dem Ziel, die Armut von Kindern in Deutschland

zu verringern, mit Hilfe von Simulationsmodellen wissenschaftlich bewerten.⁵

Eine wachsende Konzentration auf evidenzbasierte Vorabbewertungen von Reformoptionen kann allerdings dazu führen, dass der Dialog zwischen Familienpolitik und Wissenschaft kleinteilig und schnelllebig wird, da er sich am politischen Tagesgeschäft ausrichtet. Zwar gehört auch dies zu den Aufgaben einer wissenschaftlichen Politikberatung, die das Ziel einer evidenzbasierten Politikgestaltung ernstnimmt. Zugleich jedoch ist hier der inhärente Konflikt für die Beratenden mit den Zielen der Wissenschaft, die dem Prinzip des „publish or perish“ folgt, besonders virulent.

Diesen Konflikt können nicht nur die Universitäten, sondern auch die Forschungsinstitute, die diese Art der Beratung der Familienpolitik heute primär tragen, nicht leicht bewältigen. Sie stehen inzwischen ebenfalls unter immensem Druck, hochgesteckte Publikationsziele zu erreichen, um bei den Evaluationen ihrer eigenen Leistungsfähigkeit, aber auch im Hinblick auf ihre Ziele bei der Entwicklung des wissenschaftlichen Nachwuchses erfolgreich zu sein. Es zeichnet sich deswegen eine Entwicklung ab, dass sich die Institute stärker auf weniger disruptive, längerfristige Vorhaben mit hohem wissenschaftlichen Anspruch ausrichten – eine Vorgehensweise, die sich oft auch im Hinblick auf die zu erreichenden Drittmittelziele als vorteilhaft erweist.

Ein möglicher Ansatzpunkt, um unter diesen Rahmenbedingungen eine exzellente Qualität der wissenschaftlichen Beratung für die Familienpolitik zu sichern, wäre, die für eine evidenzbasierte Entscheidungsorientierung gebrauchten Modellinfrastrukturen der Institute durch die zuständigen Ressorts verlässlich zu finanzieren. Dies setzte einen Impuls, diese kontinuierlich zu pflegen und nach dem neuesten Stand der Wissenschaft weiterzuentwickeln. Konkurrierende Modelle aufrecht zu erhalten, wäre dabei im Sinne der wissenschaftlichen Qualitätssicherung, weil sich daraus Möglichkeiten zur Kreuzvalidierung und Replikation ergeben.

Ein anderer Ansatzpunkt wäre, jenseits der Detailfragen aus dem politischen Tagesgeschäft, die Wirksamkeit des Leistungskatalogs der Familienpolitik als Ganzes erneut auf den Plan zu rufen. Dies wäre ein

5 BÖHMER ET AL. 2016.

starker Impuls in Richtung Wissenschaft und beileibe keine Wiederholung des Gleichen. Der wirtschaftliche und gesellschaftliche Rahmen, in dem die Familienpolitik ihre Instrumente einsetzt, hat sich gewandelt, eine neue Generation von Eltern wächst nach, und die Methoden der Evaluation werden durch wissenschaftlichen Fortschritt besser. Deshalb ist nach der Gesamtevaluation im Idealfall vor der Gesamtevaluation.

Literatur

Böhmer, M.J., Bonin, B., Stichnoth, H. (2016). „Evaluation von Reformoptionen für eine verbesserte materielle Absicherung von Kindern. Endbericht“. Gutachten im Auftrag der Bündnis 90/Die Grünen Bundestagsfraktion. Zentrum für Europäische Wirtschaftsforschung. Mannheim.

Borchert, J. (2013). „Transferausbeutung oder Wundertüte: Welchen Beitrag leistet die Gesamtevaluation zur Familienpolitik?“ In: Der Paritätische Gesamtverband (Hrsg.), Problemfall Familienförderung – Was kommt nach der Gesamtevaluation? Dokumentation der Fachtagung vom 23. Oktober 2013, Berlin.

Deutscher Bundestag (2006). „Siebter Familienbericht: Familie zwischen Flexibilität und Verlässlichkeit – Perspektiven für eine lebenslaufbezogene Familienpolitik und Stellungnahme durch die Bundesregierung“. Unterrichtung durch die Bundesregierung, 16. Wahlperiode, Drucksache 16/1360.

Prognos (2014). „Endbericht. Gesamtevaluation der ehe- und familienbezogenen Leistungen in Deutschland“. Berlin und Basel.

Schröder, M., Siegers, R., Spieß, C.K. (2013). „Familien in Deutschland – FiD: Enhancing Research on Families in Germany“. SOEPpapers 556. Berlin: Deutsches Institut für Wirtschaftsforschung.

Zeit, Geld und Infrastruktur: Die neue Familienpolitik - Begründung und Evaluation

Von HANS BERTRAM

Korrespondenzadresse:

c/o Nationale Akademie der Wissenschaften Leopoldina

Reinhardtstr. 14

10117 Berlin

hbertram@sowi.hu-berlin.de

1. Internationale Vergleiche als Basis „neuer“ Familienpolitik, Prüfung der Effekte auf der Basis nationaler Simulationsmodelle

2003 beauftragte die damalige Familienministerin Renate Schmidt die Kommission zum Siebten Familienbericht, Konzepte und Perspektiven für eine integrative Familienpolitik zu entwerfen, um die ausufernden und teilweise nicht mehr überblickbaren Leistungen für Familien zusammenzuführen. In der Geschichte der Familienpolitik in Deutschland lassen sich vereinfachend zwei Positionen ausmachen: Zum einen sind Leistungen, die eine Familie vor allem bei der Entwicklung des Humankapitals erbringt,¹ durch einen „Leistungsausgleich“ zu stützen und zu fördern; zum anderen müssen Nachteile bei der finanziellen Leistungsfähigkeit von Familien steuerlich ausgeglichen werden. Diese eher bei konservativen Politikern dominante Sicht² wird ergänzt um die bei sozialdemokratischen Politikern vorherrschende Perspektive,³ durch die Entwicklung der Infrastruktur, vor allem den Ausbau von Krippe, Kindergarten und Ganztagschule, bestehe die Möglichkeit, dass Vater und Mutter in gleicher Weise am Arbeitsmarkt partizipieren mit der Konsequenz der Reduktion von Leistungsungleichheiten gegenüber Paaren ohne Kinder.

Der Siebte Familienbericht (2006) hat diese beiden Positionen nicht als Gegensatz betrachtet, sondern sie mit einer dritten Perspektive als integralen Bestandteil der Trias von Zeit, Geld und Infrastruktur zur Ba-

1 FÜNFTER FAMILIENBERICHT 1995.

2 BERTRAM 2017.

3 ESPING-ANDERSEN 1991.

sis einer neuen Familienpolitik konstruiert. Denn unter einer lebensverlaufstheoretischen Perspektive lässt sich zeigen, dass die Leistungen der Familie für die Gesellschaft im Lebenslauf sehr unterschiedlich verteilt sind. Zudem ist heute anders als in den 1950er und 1960er Jahren infolge der geringeren Kinderzahlen und der engen Abfolge von zweiter und dritter Geburt bei Frauen mit mehreren Kindern die Reproduktionsphase kurz geworden. Selbst bei der konservativen Vorstellung, die Kinder bis zum 15. Lebensjahr zu Hause zu betreuen, haben die Mütter dann erst das halbe Leben gelebt. Damit ist die Mutterrolle ebenso wie die Vaterrolle nur noch eine begrenzte Phase im Leben von Eltern. Diese gewonnenen Jahre in Kombination mit der heute erreichten Qualifikation von jungen Frauen und jungen Männern macht eine Familienpolitik erforderlich, die Zeit im Lebenslauf und im Alltag als Organisationsprinzip nutzt, um Infrastruktur und Geldleistungen so zu organisieren, dass sie optimale Unterstützungsleistungen im Lebenslauf bieten. Das einkommensabhängige Elterngeld, der Ausbau der Krippe und des Kindergartens und die Rechtsansprüche auf Krippe, Kindergarten und Ganztagschule sind Teile einer solchen zeitbezogenen Familienpolitik. Dahinter steht auch die Prämisse, dass die im Lebenslauf sehr unterschiedlich verteilten Fürsorgezeiten für Kinder den gleichen Stellenwert für die Entwicklung einer Gesellschaft haben wie die Tätigkeit am Markt; das erklärt die Einkommensabhängigkeit des Elterngelds sowie den Anspruch, dass sich auch die Väter aktiv daran beteiligen.

Methodisch hat der Siebte Familienbericht mit empirischen internationalen Vergleichen gearbeitet, was damals teilweise erstmalig möglich war. Die OECD baute damals ihre familienbasierte Datenbank auf, die sie bis heute auch fortschreibt, und ermöglicht damit den Vergleich der finanziellen Leistungen im OECD-Kontext.⁴ Eurostat (2017) lieferte verlässliche Vergleichsdaten zum Ausbau der Infrastruktur, und die Statistikämter Schwedens und Finnlands haben im Auftrag von Eurostat eine Datenbank aufgebaut, die die international vergleichbaren Zeitbudgetstudien so aufbereitet hatte, dass auch hier entsprechende Vergleiche möglich waren.⁵

Kurz vor Ende der Legislaturperiode hat Renate Schmidt 2005 den

4 ADEMA ET AL. 2009; OECD 2017.

5 HETUS 2018.

Siebten Familienbericht vorgestellt und die dort entwickelten Perspektiven als Vermächtnis hinterlassen. Sicherlich ungewöhnlich für einen Bericht hat Ursula von der Leyen als Nachfolgerin den Bericht erneut vorgestellt, um damit deutlich zu machen, dass sie das dort entwickelte Konzept auch politisch umzusetzen gewillt war. Es wurde eine Kommission bei der Ministerin eingerichtet, um den Umsetzungsprozess zu begleiten, die auch die Idee entwickelte, das in die Praxis umgesetzte Konzept des Siebten Familienberichts zu evaluieren, mit der schon älteren Idee des Ministeriums, die Effektivität der einzelnen Leistungen für Familien zu prüfen.

Im Gegensatz zum Siebten Familienbericht, der mit entsprechenden empirischen Daten im internationalen Vergleich gearbeitet hat, wurde bei der Evaluation entschieden,⁶ mit dem Sozio-oekonomischen Panel ein Simulationsmodell zu erstellen und die einzelnen Maßnahmen dort einzufügen, um die Effektivität der Maßnahmen im Rahmen dieses Modells zu prüfen. Zweifellos ist es die Stärke von Simulationsmodellen, komplexe Zusammenhänge anschaulich darzustellen und Einzel- und Interaktionseffekte verschiedener Maßnahmen herauszuarbeiten. Dieser Vorteil hängt auch damit zusammen, dass bestimmte Annahmen bei der Konstruktion des Modells formuliert werden, die sich selbst der direkten Überprüfung entziehen, weil sie die Konstruktion des Gesamtmodells tragen.

Die Autoren des Evaluationsmodells sind nicht von der gleichberechtigten Trias Zeit, Geld und Infrastruktur ausgegangen, sondern von der Annahme, dass familienpolitische Maßnahmen zur Integration der Mütter in den Arbeitsmarkt durch Anreizmodelle wirken. Zudem haben sie unterstellt, dass die Fürsorgezeit für Kinder letztlich „freie Zeit“ sei, die disponibel ist. Damit entspricht dieses Modell dem früheren „sozialdemokratischen Modell“ von Esping-Andersen. Dieser ging Anfang der 1990er Jahre noch davon aus, dass die sozialdemokratischen Länder durch den Ausbau der Infrastruktur und die klare gesetzliche Orientierung zur Gleichstellung von Mann und Frau auch tatsächlich zu dieser Gleichstellung am Arbeitsmarkt beitragen. Unter dieser Konstruktionsperspektive ist die abhängige Variable die Präsenz und der Zeitaufwand der Mütter am Arbeitsmarkt. In diesem Modell vermindern Geldleis-

6 BÖHMER 2014.

tungen für Familien die Präsenz am Arbeitsmarkt, weil sie einen Anreiz darstellen, nicht zu arbeiten; hingegen erhöht eine gute Infrastruktur die Präsenz am Arbeitsmarkt, weil sie die Vereinbarkeit von Beruf und Familie erleichtert.

Der größte Gegensatz zwischen dem Siebten Familienbericht und der Evaluation liegt darin, dass der Siebte Familienbericht von einer Rushhour zwischen dem 30. und 40. Lebensjahr ausging, wegen der Bedarfe von sehr kleinen Kindern und der zeitgleichen beruflichen Integration von Mann und Frau in den Arbeitsmarkt, die familienpolitisch nur partiell zu beeinflussen ist. Denn berufliche Karrieremuster, die in der gleichen Zeit die Weichenstellung für die berufliche Entwicklung setzen, nehmen keine Rücksicht auf familiäre Zeiterfordernisse. Als zweite zentrale Abweichung deutet die Evaluation die Fürsorgezeit für Kinder und ältere Menschen als „freie Zeit“ um; hingegen ging der Familienbericht davon aus, dass sich diese Zeit der Disponibilität sowohl der Eltern wie auch der Kinder, die für ihre Eltern sorgen, entzieht.

Hier liegt eine in Wissenschaft und Politikberatung seltene Situation vor, dass zwei verschiedene theoretische Perspektiven und konträres empirisches Vorgehen die Prüfung ermöglichen, ob die Ergebnisse der Simulation im Evaluationsmodell einem empirischen Realitätstest standhalten und ob die geänderten theoretischen Annahmen des Familienberichts letztlich zu falschen oder richtigen Empfehlungen für die Politik geführt haben. Diese Prüfung kann hier nicht im Detail erfolgen, sondern nur an wenigen Beispielen. Die These des Familienberichts, die Fürsorgezeit sei keine freie Zeit, sondern beanspruche im Lebenslauf festliegende Zeiten, die gegenwärtig eine extreme Rushhour hervorrufen, ist mit den aufbereiteten Zeitbudgetstudien des schwedischen und finnischen Statistikamts für ganz Europa zu prüfen. Die Grundannahme, vor allem staatliche finanzielle Anreize beeinflussten das Erwerbsverhalten von Müttern, lässt sich durch Eurostat und den Mikrozensus prüfen.

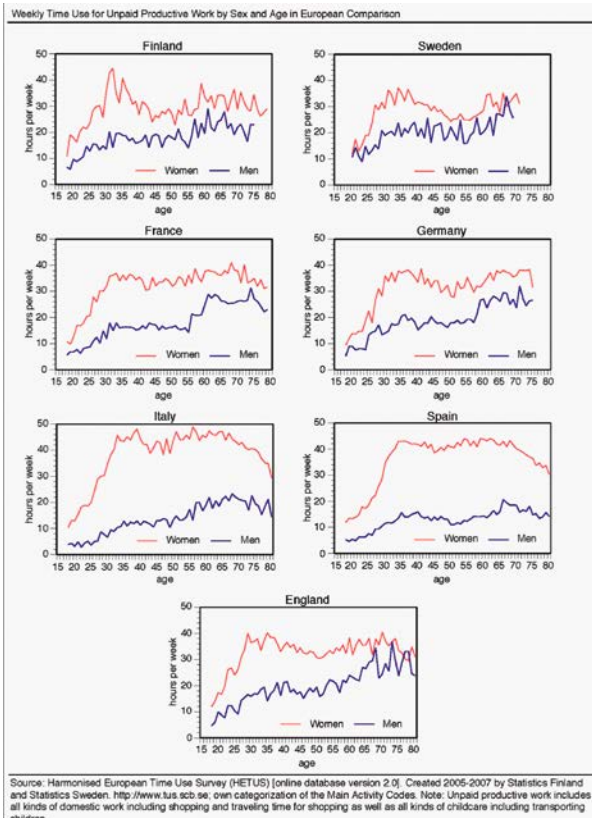
2. Rushhour im Lebensverlauf im internationalen Vergleich

Die nordeuropäischen Staaten⁷ gelten immer als Vorbild für eine gute Infrastruktur für Kinder aller Altersstufen und ihre Gesetzgebung wird

7 ESPING-ANDERSEN 2016; GOLDSCHNEIDER ET AL. 2015.

in der Regel als gleichstellungsorientiert interpretiert.⁸ All die Zeit, die im Haushalt mit Kindern sowie mit Haushaltsführung, Einkaufen und Kochen verbracht wird und die sich grundsätzlich auch über den Markt einkaufen ließe, wird als produktive Tätigkeit bezeichnet, die nicht bezahlt wird: Die Daten der harmonisierten Zeitbudgetstudien (Hetus) zeigen diese Zeit im Lebensverlauf von Männern und Frauen für Finnland, Schweden, Frankreich, Deutschland, Italien, Spanien und England im Vergleich (vgl. Grafik).

Abbildung 1: Zeitverwendung für unbezahlte produktive Tätigkeiten nach Alter und Geschlecht



8 ESPING-ANDERSEN, BILLARI 2015.

In Finnland, Schweden, Deutschland und nicht ganz so deutlich auch in England ist ein ausgeprägter Zeitgipfel zwischen dem 28. und 40. bis 45. Lebensjahr zu beobachten, der in Deutschland 37 bis 38 Stunden ausmacht, in Schweden etwa 36 Stunden und in Finnland etwas über 40 Stunden, ähnlich wie in England; der Kurvenverlauf ist in Abhängigkeit vom Lebensverlauf deutlich erkennbar. In Italien, Spanien und Frankreich gibt es einen kontinuierlicheren Verlauf, wobei Frankreich mit 40 Stunden weniger Zeitaufwand ausweist als Italien mit fast 50 und Spanien mit etwa 45 Stunden. Die männliche Zeit für solche Tätigkeiten während der Rushhour-Phase zwischen dem 30. und 40. Lebensjahr liegt in Finnland bei etwa 20 Stunden, Schweden hat 22 bis 23 Stunden, Deutschland 20 Stunden und Frankreich 15 bis 18 Stunden; in Spanien und Italien sind es etwas mehr als 10 Stunden und in England knapp 20 Stunden. Allerdings steigt der Zeitaufwand der Männer im Haushalt im Lebensverlauf deutlich an.

Frankreich mit seiner hervorragenden Infrastruktur für Kinder unter 3 und unter 6 Jahren weist genauso wie Schweden oder Finnland einen Anstieg des Zeitaufwands für die Haushaltstätigkeiten auf und unterscheidet sich hier nicht von Deutschland. Entscheidende Unterschiede in den Kurven zeigen sich erst nach dem 40. Lebensjahr, wenn in Finnland der Zeitaufwand deutlich zurückgeht und in Schweden nicht ganz so deutlich. Deutschland folgt dem schwedischen Modell mit einem weniger deutlichen Rückgang. Frankreich folgt mit dem geringeren Zeitaufwand eher dem südeuropäischen Muster wie Italien und Spanien.

Nach diesen Zeitverläufen hat die vorhandene Infrastruktur auf den Zeitaufwand für Kinder nur einen begrenzten Einfluss. Wenn die Kinder sehr klein sind, fällt in allen Ländern im Durchschnitt aller Frauen eine Arbeitszeit von etwa 40 Stunden an. Mit dem Lebensalter der Kinder als Grundgesamtheit und nur den bis zu 3 bzw. bis zu 7 Jahren alten Kindern ist die zeitliche Belastung für diese Gruppe in allen Ländern noch ausgeprägter und liegt zwischen 50 und 60 Stunden. Die Vorstellung der Evaluationsgruppe und vieler ökonomischer Theorien (Backups), diese Hausarbeitzeit sei „frei disponibel“, ist mit diesen Zeitbudgetdaten falsifiziert.

Wer sich für Kinder entscheidet, muss unabhängig von den staatlichen Maßnahmen und politischen Strategien damit rechnen, zwischen dem 30. und 45. Lebensjahr in eine extreme zeitliche Belastung

zu kommen. Die Rushhour des Lebens lässt sich als Zeitphase nicht so einfach staatlich beeinflussen, sondern hängt offensichtlich viel mehr von den Bedürfnissen der Kinder, von kulturellen Kontexten (Spanien, Italien, Frankreich) und den Unterstützungsleistungen der Männer ab. Die Unterstützungsleistung der deutschen Männer entspricht etwa der der finnischen Männer mit rund 20 Stunden. Daher können Modelle, die durch die Angleichung der Hausarbeit von Mann und Frau die Präsenz der Frauen am Arbeitsmarkt zu erhöhen hoffen, das nur dadurch erreichen, dass die zeitliche Belastung der Eltern mit Kindern insgesamt immer deutlich höher liegt als für diejenigen, die sich für andere Lebenswege entschieden haben.⁹

Diese strukturelle Rushhour des Lebens falsifiziert das Anreizmodell des Evaluationsberichts. Denn auch in Ländern, in denen die Vereinbarkeit politisch gewollt und institutionell auch weitgehend umgesetzt ist, etwa Schweden, entsprechen die Zeitmuster den deutschen Zahlen. Inzwischen liegen Zeitbudgetdaten auch für 2011/2012 (Statistisches Bundesamt 2015) vor, so dass sich prüfen lässt, ob sich seit 2001 viel geändert hat, was nicht Fakt ist.¹⁰ Der Ausbau der Infrastruktur in Deutschland hatte und hat noch einen ähnlichen Effekt wie in Finnland: 2005 gaben 60 Prozent der Eltern in Deutschland an, ihre unter 3-jährigen Kinder allein zu betreuen, und zudem 40 Prozent, dass ihre Kinder noch durch andere Personen betreut werden.¹¹ Diese Daten entsprechen ziemlich den Zahlen aus Finnland.

Auch 2015 hat sich diese Zahl nicht verändert, so dass die Frage im Raum steht, wer die vielen jetzt angebotenen Krippenplätze nachfragt. Mit Hilfe von Eurostat ist das gut zu klären: 40 Prozent Eltern, die sich schon vor der neuen Familienpolitik dafür entschieden hatten, frühzeitig in den Beruf zurückzukehren, haben ihr Verhalten massiv geändert und nutzen zunehmend die kommunalen und staatlichen Angebote zur Kinderbetreuung. Die Relation 40 zu 60 bedeutet auch, dass man rund 40 Prozent, in manchen Großstädten 50 Prozent Plätze für Kinder unter 3 Jahren anbieten muss, weil die Verlässlichkeit und Qualität staatlicher

9 ZAGHENI ET AL. 2015; ZAGHENI, ZANELLA; BERTRAM 2017.

10 BERTRAM 2017.

11 BERTRAM 2017.

Betreuungsangebote von den Eltern in Deutschland geschätzt wird.¹²

In diesen Zahlen verbirgt sich auch ein Lösungsmodell mit der Rushhour des Lebens. Die Gruppe, die ihre Kinder bis zum 3. und möglicherweise bis zum 6. Lebensjahr im Wesentlichen allein betreut, folgt bei der Integration ins Berufsleben dem sequenziellen Modell, wie es in den 1930er und 1940er Jahren Myrdal und Klein angedacht hatten. Die anderen Eltern folgen einem parallelen Modell, bei dem die Arbeitszeit während der Rushhour des Lebens in allen Ländern deutlich reduziert wird, um danach wieder vollzeitnah zu arbeiten. Das Problem des Evaluationsmodells ist, dass das sequenzielle Modell als Lösungsmodell der Rushhour überhaupt nicht bedacht wurde.

Ob die deutschen Eltern in Zukunft weiterhin dem finnischen Beispiel folgen oder dem schwedischen, ist offen, weil sich keine Prognosen über das zukünftige Verhalten machen lassen. Trotzdem lassen sich relativ klare politische Aussagen formulieren. Solange der Zeitaufwand für die Fürsorge für Kinder zwischen Männern und Frauen asymmetrisch verteilt ist - und das in allen Ländern, die mit den Zeitbudgetdaten dargestellt wurden - ist die Gleichheit von Einkommen und Karriere am Arbeitsmarkt nicht erreichbar. Die OECD hat Schweden wegen des Gender Pay Gaps massiv kritisiert,¹³ der sich nach den OECD-Daten kaum von Deutschland unterscheidet.¹⁴ In Deutschland ist der Gender Pay Gap nach den OECD-Daten eher ein Mütter Pay Gap.

3. Eine evidenzbasierte Kritik des Simulationsmodells zur Prüfung der nachhaltigen Familienpolitik

Die Frage, ob die Fürsorgezeit für Kinder als „freie Zeit“ zu interpretieren ist als Kernannahme des Simulationsmodells, wurde unter einer lebenslauftheoretischen Perspektive falsifiziert. In keinem der untersuchten Länder konnte auch bei ausgebauter Infrastruktur und entsprechender Familienpolitik die Rushhour im Lebensverlauf von Müttern zwischen dem 30. und 40. Lebensjahr beseitigt werden. Die Gründe dafür wurden hier nicht untersucht. Möglich sind verschiedene Präferenzen von

12 BERGER 2017.

13 OECD 2012.

14 OECD 2017.

Müttern und Vätern,¹⁵ anthropologische Dispositionen,¹⁶ kulturelle Überlieferungen,¹⁷ unterschiedliche Einstellungen zur Arbeitswelt¹⁸ und anderes. Das war auch nicht notwendig, weil allein der Nachweis, dass zu einem bestimmten Zeitpunkt eine bestimmte zeitliche Belastung auftritt, zeigt, dass gegenwärtig unabhängig von der jeweiligen Familienpolitik mit einer Rushhour im Leben zu rechnen ist.

Das Konzept der evidenzbasierten Medizin ist inzwischen auch in die soziale Arbeit übertragen worden, bisher weniger in Deutschland als in den USA, Großbritannien und den skandinavischen Ländern.¹⁹ Diese Konzepte basieren auf drei zentralen Elementen. Soziale, pädagogische oder therapeutische Maßnahmen für betroffene Klienten haben die Präferenzen der Betroffenen zu berücksichtigen sowie ihre Lebenssituation und ihre Perspektiven in die jeweilige Konzeption der Behandlung einzubeziehen. Solche an den Interessen der Betroffenen orientierten Konzepte sollen abgeleitet werden aus den besten, systematisch aufbereiteten Forschungen zum jeweiligen Themengebiet. Und an diese systematische Aufbereitung der Forschung ist die Forderung gestellt, dass die Forschungsergebnisse eindeutige Hinweise auf die intendierte Wirkung der geplanten Maßnahmen geben.

Dieses Modell wird in Deutschland in der sozialen Arbeit und der Familienpolitik noch recht selten praktiziert, doch lässt sich der Anspruch der nachhaltigen Familienpolitik schon dahingehend interpretieren, dass die persönlichen und subjektiven Präferenzen der betroffenen Mütter und Väter zentrales Grundelement einer solchen Politik sein müssen, die vorgeschlagenen Maßnahmen und gesetzlichen Regelungen auch im internationalen Vergleich durch systematische Forschung geprüft worden sind und davon auszugehen ist, dass die intendierten Wirkungen auch eintreten werden. Nach diesen Kriterien ist das Simulationsmodell nicht als ein evidenzbasiertes Modell zu bezeichnen. Denn die Diskussion zur Kernannahme des Simulationsmodells, ob die Fürsorgezeit für Kinder als freie Zeit zu interpretieren ist, hat gezeigt, dass

15 MOEN ET AL. 2012.

16 HRDY 2012.

17 COONTZ 2012.

18 GILLIS 2012.

19 DAHMEN 2011.

diese Annahme unter einer lebenslauftheoretischen Perspektive als falsifiziert gilt, weil in keinem der untersuchten Länder die Rushhour im Lebensverlauf von Müttern beseitigt werden konnte. In allen Ländern wird in dieser Rushhour entweder die Arbeitszeit deutlich reduziert, auch in den nordeuropäischen Ländern, oder ganz auf die Integration in den Arbeitsmarkt verzichtet.

Dahinter verbergen sich das parallele und sequenzielle Modell. Das Modell, Familienarbeit und Berufsarbeit parallel im Lebensverlauf zu gestalten, führt zu einer eher kurzen Unterbrechung für Fürsorgeleistung für die Kinder und zur Reduktion der Arbeitszeit, in der Regel bis etwa zum 8. Lebensjahr des Kindes. In Finnland und in Deutschland dominiert mit jeweils 60 Prozent der betroffenen Mütter das sequenzielle Modell, auch wenn das politisch nicht gewollt sein mag und man auch als Wissenschaftler plausible Gründe gegen ein solches Modell vortragen kann.²⁰ Bei einer evidenzbasierten Perspektive, die die Interessen, Wünsche und Lebensvorstellungen der Betroffenen berücksichtigt, ist das sequenzielle Modell aber ebenso zu akzeptieren wie das parallele Modell. Wie es der Siebte Familienbericht gesehen hat, muss eine nachhaltige Familienpolitik die Entscheidung für die Gestaltung des Lebenslaufs dem Einzelnen überlassen und die Bedingungen so konzipieren, dass parallele und sequenzielle Modelle gelebt werden können.

Die Annahme, dass es keine Rushhour im Leben gäbe, weil diese als „freie Zeit“ beliebig im Lebensverlauf zu disponieren sei, führt dazu, das sequenzielle Modell zugunsten des Parallelmodells aufzugeben. Aus politischen und wirtschaftlichen Gründen mag das sinnvoll sein. Das ist dann aber kein evidenzbasiertes Modell mehr, weil hier die Wünsche und Lebensvorstellungen einer großen Gruppe, nämlich momentan 60 Prozent der Betroffenen, ausgeblendet werden. Ein evidenzbasiertes Modell zur Prüfung nachhaltiger Familienpolitik muss so konzipiert sein, dass auf der einen Seite die empirische Realität (Rushhour des Lebens) nicht ausgeblendet wird und zugleich die verschiedenen Präferenzen, mit dieser Rushhour umzugehen, berücksichtigt sind.

Das Modell ist auch deswegen nicht als evidenzbasiert zu bezeichnen, weil es bei seiner Konstruktion nicht die schon damals vorliegenden aufbereiteten Zeitbudgetdaten (Hetus) integriert hat. Vermutlich

20 MOEN ET AL. 2012.

haben die Konstrukteure des Simulationsmodells eine weitere Annahme empirisch nicht überprüft. Sie gehen davon aus, dass staatliche ökonomische Anreize das Verhalten der Mütter und Väter wesentlich beeinflussen, und dass diese Anreize relativ unabhängig vom Kontext überall die gleiche Wirkung entfalten. Der Mikrozensus weist aus,²¹ dass 2015 im Freistaat Bayern 75 Prozent der 25- bis 49-jährigen Frauen nach eigenen Angaben vom eigenen Einkommen leben; 20 Prozent leben vom Ehemann und 5 Prozent von staatlichen Unterstützungsleistungen. In den neuen Bundesländern leben nach eigenen Angaben in der gleichen Altersgruppe auch 75 Prozent vom eigenen Einkommen, aber nur 5 Prozent vom Ehemann und 20 Prozent von staatlicher Unterstützung. In den neuen Bundesländern ist der Staat an die Stelle des Ehemanns getreten. Dabei unterscheiden sich die bayerischen Zahlen nicht von den nordeuropäischen oder französischen Zahlen.

Mit diesen Zahlen ist zunächst auszuschließen, dass steuerliche Anreize, wie das Splitting-System, überall die gleiche Wirkung entfalten. Denn beim gleichen Steuersystem verhalten sich die 25- bis 49-jährigen Frauen in Bayern konträr zu den Altersgenossinnen in den neuen Bundesländern, wenn sie nicht vom eigenen Einkommen leben können: Die einen stützen sich auf den Ehemann, die anderen auf den Staat. Auch ein internationaler Vergleich zeigt deutlich, dass die Präsenz am Arbeitsmarkt mit der steuerlichen Belastung und Konstruktion des Steuerrechts kaum in Beziehung zu setzen ist. Frankreich mit einem ähnlichen Splitting-System wie Deutschland zeigt seit langem eine hohe Präsenz der Frauen und Mütter am Arbeitsmarkt, während das in Deutschland erst seit den 1990er Jahren zu beobachten ist.

In einem evidenzbasierten Modell ist es erforderlich, diese systematischen Abweichungen im nationalen wie im internationalen Kontext zu berücksichtigen und sie auch als Randbedingungen in ein entsprechendes Simulationsmodell zu integrieren. Sie im Modell gar nicht zu berücksichtigen, bedeutet schlicht, dass mögliche vermutete Wirkungen im Simulationsmodell nur dadurch Evidenz gewinnen können, dass abweichende Ergebnisse ignoriert werden. Genau das aber sollte durch das Konzept der Evidenz in der Medizin wie in der sozialen Arbeit überwunden werden, denn viel zu lange wurden in den beteiligten Fächern

21 STATISTISCHES BUNDESAMT 2015; BERTRAM 2017.

nur die Ergebnisse berücksichtigt, die in das eigene Weltbild und die eigene Ergebnisstruktur passten. Evidenzbasierung erfordert aber gerade, alle relevante und für das Themengebiet wichtige Literatur und die dafür zentralen Daten und Ergebnisse zu berücksichtigen.

Damit ist aber auch der dritte Aspekt evidenzbasierter Familienpolitik, einen klaren Wirkungszusammenhang herzustellen, nicht gewährleistet. Denn selbst wenn das Simulationsmodell die aktuelle Wirklichkeit abbildet, ist es nicht in der Lage zu erklären, ob die Ursachen bestimmter Entwicklungen Folge einer bestimmten Politik sind oder andere Ursachen haben, etwa langfristige Trends. So wurde die Integration der jungen Frauen in die Arbeitswelt im Wesentlichen während der Regierungszeit von Helmut Kohl vollzogen; die Steigerungsraten seit 2000 schwanken kontinuierlich um 1 bis 2 Prozent. Ob diese Schwankungen nun auf die neue Familienpolitik zurückzuführen sind oder „nur“ Ergebnis der sich langsam wandelnden Orientierung der nachwachsenden Frauengeneration ist, lässt sich durch das Simulationsmodell nicht klären. Das ist bedauerlich, weil solche Modelle nun Gefahr laufen, nur noch zur Legitimation politischer Maßnahmen herangezogen zu werden. Dieser Gefahr sollte die Wissenschaft nicht erliegen.

Diese skeptische Beurteilung des Simulationsmodells zur Überprüfung nachhaltiger Familienpolitik bedeutet nicht, solche Modelle generell oder auch dieses Modell zu verwerfen. Die Modelle haben den Vorteil, die Annahmen verändern zu können, so dass sich möglicherweise doch bestimmte Zusammenhänge nachhaltiger Familienpolitik überprüfen lassen. Auch haben sie den Vorteil, dass sich ihre Ergebnisse als empirisch zu prüfende Hypothesen betrachten lassen, die bestätigt oder falsifiziert werden können.

Der Nachweis, dass neben dem parallelen Modell, Beruf und Familie zu vereinbaren, auch ein sequenzielles Modell existiert, in Europa besonders ausgeprägt in Finnland und Deutschland, hat bestimmte politische Schlussfolgerungen. Das sequenzielle Modell führt durch die längere Unterbrechung bei den gegenwärtigen Karrieresystemen notwendigerweise zu einer systematischen Benachteiligung der Mütter in ihrer beruflichen Entwicklung und beim Einkommen. Der Siebte Familienbericht hatte angesichts dieser Tatsache vorgeschlagen, dynamische Lebenslaufmodelle zu entwickeln, die nicht mehr an eine feste Lebensarbeitszeit und ein festes Austrittsalter gebunden sind, sondern

entsprechend der verschiedenen Modelle parallel oder sequenziell so konzipiert werden, dass gleiche Ansprüche in unterschiedlichen Zeiträumen erworben werden können. Auf diese Weise werden die gewonnenen Jahre genutzt, um unterschiedliche Lebensentwürfe leben zu können und trotzdem ein gleiches Einkommen und eine gleiche Rente zu erzielen.

Diese Perspektive steht ziemlich konträr zur gegenwärtigen Alterssicherungspolitik, die an starren Grenzen festhält. Unter einer familien- und frauenpolitischen Perspektive ist aber davon auszugehen, dass die sich daraus ergebende Benachteiligung bestehen bleibt, solange viele Mütter ein sequenzielles Modell wählen, obwohl das über eine Flexibilisierung der Altersperspektiven aufzulösen wäre. Abschließend sei nicht ohne Interesse vermerkt, dass dieser Teil des Siebten Familienberichts weder von der Politik noch von den Autoren des Simulationsmodells zur Überprüfung einer nachhaltigen Familienpolitik aufgegriffen wurde.

Literatur

Adema, W. et al. (2009). „The OECD Family Database: Developing a Cross-National Tool for Assessing Family Policies and Outcomes“. *Child Indicators Research* (2):4, 437–60, doi:10.1007/s12187-009-9044-8.

Berger, H. (2017). „Nachgefragt: Was denken Eltern über KiTas?“. Gütersloh: Bertelsmann doi:10.11586/2018002.

Bertram, H. (2017). „Die Zweiverdiener-Familie: Ein europäischer Vergleich“. Berlin: Konrad-Adenauer-Stiftung.

Böhmer, M. et al. (2014). „Gesamtevaluation der ehe- und familienbezogenen Maßnahmen und Leistungen in Deutschland. Endbericht“. Berlin: Prognos.

Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.) (1995). „Fünfter Familienbericht: Familien und Familienpolitik im geeinten Deutschland. Zukunft des Humanvermögens“. Verhandlungen des Deutschen Bundestages. Bonn.

Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg.) (2006). „Siebter Familienbericht. Familie zwischen Flexibilität und Verlässlichkeit: Perspektiven für eine lebenslaufbezogene Familienpolitik“. Verhandlungen des Deutschen Bundestages. Berlin.

- Coontz, S. (2012). „The Late Arrival and Early Demise of the Male Breadwinner Family”. In: Bertram, H., Ehlert, N. (Hrsg.), *Family, Ties, and Care. Family Transformation in a Plural Modernity*. Opladen: Barbara Budrich
- Dahmen, S. (2011). „Evidenzbasierte soziale Arbeit? Zur Rolle wissenschaftlichen Wissens für sozialarbeiterisches Handeln”. Hohengehren (Baltmannsweiler): Schneider.
- Esping-Andersen, G. (1991). „The Three Worlds of Welfare Capitalism”. Cambridge.
- Esping-Andersen, G. (2016). „Families in the 21st Century”. SNS Förlag.
- Esping-Andersen, G., Billari, F.C. (2015). „Re-Theorizing Family Demographics”. *Population and Development Review* 41:1, 1–31, doi:10.1111/j.1728-4457.2015.00024.x.
- Eurostat (2017). „Final Report of the Quality of Life Indicators”. Brüssel.
- Eurostat (2017). „Glossar: EU-Statistik über Einkommen und Lebensbedingungen (EU-SILC) – Statistics Explained”. [http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_statistics_on_income_and_living_conditions_\(EU-SILC\)/de](http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:EU_statistics_on_income_and_living_conditions_(EU-SILC)/de) [Zugriff 12. April 2017]
- Gillis, J.R. (2012). „Marginalization of Fatherhood in Western Countries”. In: Bertram, H., Ehlert, N. (Hrsg.), *Family, Ties, and Care. Family Transformation in a Plural Modernity*. Opladen: Barbara Budrich, 85ff.
- Goldscheider, F. et al. (2015). „The Gender Revolution: A Framework for Understanding Changing Family and Demographic Behavior”. *Population and Development Review* 41:2, 207–39, doi:10.1111/j.1728-4457.2015.00045.x.
- HETUS (2018). „HETUS - Harmonised European Time Use Survey”. <https://www.h6.scb.se/tus/tus/>. [Zugriff 30. Juli 2018]
- Hrdy, S.B. (2012). „Cooperative Breeding and the Paradox of Facultative Fathering”. In: Bertram, H. & Ehlert, N. (Hrsg.), *Family, Ties, and Care. Family Transformation in a Plural Modernity*. Opladen: Barbara Budrich, 207ff.
- Moen, P., Kelly, E., Magennis, R. (2012). „Gender Strategies: Socialization, Allocation, and Strategic Selection Processes: Shaping the Gendered Adult Course”. In: Bertram, H. & Ehlert, N. (Hrsg.), *Family, Ties, and Care. Family Transformation in a Plural Modernity*. Opladen: Barbara Budrich, 135ff.

OECD (2012). „Closing the Gender Gap ACT NOW – Sweden“. <https://www.oecd.org/sweden/Closing%20the%20Gender%20Gap%20-%20Sweden%20FINAL.pdf> [Zugriff: 30. Juli 2018]

OECD (2017). „Gender Wage Gap, Total, % of Male Median Wage, 2017 or latest available“. <https://data.oecd.org/chart/5i7W>

OECD (2017). „OECD Family Database“. <http://www.oecd.org/els/family/database.htm>. [Zugriff 7. Mai 2017]

Statistisches Bundesamt (2015). „Zeitverwendungserhebung; Aktivitäten in Stunden und Minuten für ausgewählte Personengruppen. Sonderauswertung“. https://www.destatis.de/DE/Publikationen/Thematisch/EinkommenKonsumLebensbedingungen/Zeitbudgeterhebung/Zeitverwendung5639102139004.pdf?__blob=publicationFile [Zugriff 19. September 2018]

Zagheni, E. et al. (2015). „A Comparative Analysis of European Time Transfer between Generations“. Wiesbaden: Springer.

Zagheni, E., Zannella, M. (2013). „The Life Cycle Dimension of Time Transfers in Europe“. *Demographic Research*, doi:10.4054/Dem-Res.2013.29.35.

6. Anhang

Workshop zum Thema „Evaluierung von Finanzmarktreformen: Lehren aus anderen Politikbereichen und Disziplinen“

29. Mai 2018

Veranstaltungsort:

Nationale Akademie der Wissenschaften Leopoldina

Jägerberg 1, 06108 Halle (Saale)

9:15 – 9:25 | Begrüßung

Prof. Dr. Jörg Hacker

Nationale Akademie der Wissenschaften Leopoldina

Politikfeld Arbeitsmarkt

9:25 – 9:40

Prof. Dr. Christoph M. Schmidt

RWI - Leibniz-Institut für Wirtschaftsforschung

9:40 – 10:45

Prof. Bernd Fitzenberger, Ph.D.

Humboldt-Universität zu Berlin

Prof. Axel Börsch-Supan, Ph.D.

Munich Center for the Economics of Aging (MEA)

Prof. Dr. Ulrich Walwei

Institut für Arbeitsmarkt- und Berufsforschung

Moderation:

Dr. Johannes Ludewig

Nationaler Normenkontrollrat

10:45 – 11:00 | Gute Evidenz für wirksame Fiskalpolitik

Dr. Ludger Schuknecht

Bundesministerium der Finanzen

11:00 – 11:20 | Kaffeepause

Politikfeld Finanzmärkte

- 11:20 – 11:35** **Prof. Jörg Rocholl, Ph.D.**
ESMT Berlin
- 11:35 – 12:40** **Prof. Dr. Monika Schnitzer**
Ludwig-Maximilians-Universität München
- Dr. Thorsten Freihube**
Bundesanstalt für Finanzdienstleistungsaufsicht
- Carsten Frank**
Bundesministerium der Finanzen
- Moderation:** **Prof. Dr. Claudia M. Buch**
Deutsche Bundesbank

12:40 – 14:00 | Mittagessen

Politikfeld Gesundheit

- 14:00 – 14:15** **Prof. Dr. J.-Matthias Graf von der Schulenburg**
Gottfried Wilhelm Leibniz Universität Hannover
- 14:15 – 15:20** **Prof. Dr. Joachim Winter**
Ludwig-Maximilians-Universität München
- Prof. Dr. Reiner Leidl**
Ludwig-Maximilians-Universität München
- Prof. Dr. Gerd Gigerenzer**
Harding-Zentrum für Risikokompetenz am Max-Planck-Institut für Bildungsforschung
- Moderation:** **Prof. Dr. Martin Lohse**
Nationale Akademie der Wissenschaften Leopoldina

15:20 – 15:50 | Kaffeepause

Politikfeld Familie

15:50 – 16:05 **Prof. Dr. C. Katharina Spieß**
DIW Berlin, Freie Universität Berlin

16:05 – 17:10 **Prof. Dr. Holger Bonin**
Institut zur Zukunft der Arbeit Bonn
Petra Mackroth
Bundesministerium für Familie, Senioren,
Frauen und Jugend
Malte Ristau-Winkler
Prognos AG

Moderation: **Prof. em. Dr. Hans Bertram**
Humboldt-Universität zu Berlin

17:10 – 17:30 | Schlussworte

Prof. Dr. Claudia M. Buch
Deutsche Bundesbank

Prof. Regina T. Riphahn, Ph.D.
Nationale Akademie der Wissenschaften Leopoldina

Leopoldina-Forum

In der Reihe „Leopoldina-Forum“ dokumentiert die Nationale Akademie der Wissenschaften vielfältige Perspektiven auf Fragestellungen hoher Relevanz für Wissenschaft und Gesellschaft.

Nr. 1: Buch CM, Riphahn RT (Hrsg.) (2019): Evaluierung von Finanzmarktreformen - Lehren aus den Politikfeldern Arbeitsmarkt, Gesundheit und Familie. Nationale Akademie der Wissenschaften Leopoldina, Halle (Saale).

**Deutsche Akademie der Naturforscher Leopoldina e.V.
– Nationale Akademie der Wissenschaften –**

Deutsche Bundesbank

Jägerberg 1
06108 Halle (Saale)
Tel.: (0345) 472 39-600
Fax: (0345) 472 39-919
E-Mail: leopoldina@leopoldina.org

Postfach 10 06 02
60006 Frankfurt am Main
Besucher- und Lieferanschrift:
Wilhelm-Epstein-Straße 14
60431 Frankfurt am Main

Berliner Büro:
Reinhardtstraße 14
10117 Berlin

Tel: 069 9566-0
Fax: 069 9566-3077
info@bundesbank.de

Die **Leopoldina** wurde 1652 gegründet und versammelt mit etwa 1500 Mitgliedern hervorragende Wissenschaftlerinnen und Wissenschaftler aus rund 30 Ländern. Sie ist der freien Wissenschaft zum Wohle der Menschen und der Gestaltung der Zukunft verpflichtet. Als Nationale Akademie Deutschlands vertritt die Leopoldina die deutsche Wissenschaft in internationalen Gremien und nimmt zu wissenschaftlichen Grundlagen politischer und gesellschaftlicher Fragen unabhängig Stellung. Hierzu erarbeitet sie unabhängige Expertisen von nationaler und internationaler Bedeutung. Die Leopoldina fördert die wissenschaftliche und öffentliche Diskussion, sie unterstützt wissenschaftlichen Nachwuchs, verleiht Auszeichnungen, führt Forschungsprojekte durch und setzt sich für die Wahrung der Menschenrechte verfolgter Wissenschaftler ein.

Die **Deutsche Bundesbank** ist die Notenbank für Deutschland. Als Teil des Eurosystems leistet sie einen wichtigen Beitrag zur Stabilität der gemeinsamen Währung, des Euro. Über die Geldpolitik hinaus erfüllt die Bundesbank wichtige Aufgaben im nationalen und internationalen Umfeld. Dazu gehört die nationale Aufsicht über Kreditinstitute, auch im Rahmen der einheitlichen europäischen Bankenaufsicht, sowie die Bereiche Bargeld und unbarer Zahlungsverkehr, Finanz- und Währungsstabilität. Die Bundesbank arbeitet in allen internationalen Institutionen und Gremien mit, die der Stabilisierung des Finanzsystems verpflichtet sind. Sie verwaltet zudem Deutschlands Währungsreserven, dient als Hausbank des Staates und erfüllt wichtige Aufgaben in der Statistik. Auch berät sie die Bundesregierung in währungspolitischen Fragen.

www.leopoldina.org