

Feldexperimente für die Armutsbekämpfung: Zum Nobelpreis für Abhijit Banerjee, Esther Duflo und Michael Kremer

Verfasser: Stefan Klonner

Am 14. Oktober gab die Königlich Schwedische Akademie der Wissenschaften bekannt, dass der Wirtschaftsnobelpreis 2019 zu gleichen Teilen an die Entwicklungsländerforscher Abhijit Banerjee, Esther Duflo und Michael Kremer verliehen wird.

Damit wurden zum ersten Mal in seiner 50-jährigen Geschichte Ökonomen, die sich vorrangig mit Armut und deren Bekämpfung beschäftigen, ausgezeichnet. Im Jahr 1998 wurde der Preis an den indischen Ökonomen Amartya Sen für seine Leistungen auf dem Gebiet der theoretischen Wohlfahrtsökonomik verliehen, wobei das Preiskomitee auch seine Arbeiten über Armutsmessung und Hungersnöte erwähnte. Beim Preisträger des Jahres 2014, Angus Deaton, erwähnte das Preiskomitee ebenfalls seine Analysen zur Armut, neben seinen einflussreichen Arbeiten zu Konsum und Wohlfahrtsmessung.

Die diesjährigen Preisträger werden „für ihren experimentellen Forschungsansatz zur globalen Armutsreduzierung“ ausgezeichnet. In einer Pressemitteilung führt das Preiskomitee aus, dass die Wissenschaftler einen neuen Ansatz entwickelt haben, um belastbare empirische Evidenz zu den geeignetsten Möglichkeiten, die globale Armut zu reduzieren, zu generieren. Ihr Forschungsprozess beinhaltet, dass die große Frage der Armutsreduzierung zunächst in kleinere, überschaubare Fragen unterteilt wird und daraus wirksame Interventionen und Politikmaßnahmen entwickelt werden, um beispielsweise Gesundheit und Schulbildung von Kindern zu verbessern. Die drei Ökonomen haben gezeigt, wie diese spezifischen Forschungsfragen mit sogenannten Experimenten unter Menschen, die selbst von Armut betroffen sind, glaubhaft beantwortet werden können.

Der experimentelle Forschungsansatz in den empirischen Sozialwissenschaften

Eine grundlegende Frage in jeglicher empirischer Forschung, gleichermaßen in Naturwissenschaften, Medizin und Wirtschafts- und Sozialwissenschaften, betrifft die Wirkung einer bestimmten Maßnahme oder eines Eingriffs, während gleichzeitig alle anderen Faktoren, die die Ergebnisvariable von Interesse beeinflussen könnten, unverändert bleiben. Eine solche Wirkung bezeichnet man als den *kausalen Effekt* einer Maßnahme. Im Bereich der Medizin kann dies die Verabreichung einer geeigneten Tablette gegen Kopfschmerzen sein, im Bereich der Armutsreduzierung eine Grundrente und ihre Auswirkung auf Ernährungsstatus und Bildung von Kindern aus benachteiligten Verhältnissen.

Kausale Effekte in den Naturwissenschaften werden üblicherweise mittels sogenannter *kontrollierter Experimente* nachgewiesen. Ein wesentlicher Bestandteil solcher Experimente ist der Vergleich einer behandelten Gruppe mit einer Kontrollgruppe. In einem Tierversuch beispielsweise werden von 50 genetisch identischen Jungmäusen zufällig 25 ausgewählt und mit einem Wachstumshormon behandelt. Nach einer Woche vergleicht der Forscher die durchschnittliche Körpergröße der behandelten Tiere mit derjenigen der nicht behandelten Tiere. Die Differenz gibt dann den kausalen durchschnittlichen Effekt des Wachstumshormons auf die Körpergröße wieder. Hierbei sind drei Aspekte des Forschungsdesigns bedeutend. Erstens werden die zu behandelnden Tiere zufällig ausgewählt. Zweitens sind alle vom Forscher kontrollierbaren Faktoren mit

Ausnahme der Hormongabe identisch für die behandelten und unbehandelten Tiere. Diese beiden Verfahrensschritte stellen sicher, dass Größendifferenzen zwischen behandelten und unbehandelten Tieren am Ende des Experiments nicht von anderen systematischen Unterschieden zwischen den beiden Gruppen verursacht sind. Drittens umfassen beide Gruppen eine hinreichend große Zahl von Tieren. Dieser Aspekt erlaubt es, davon auszugehen, dass vom Forscher nicht kontrollierbare, zufällige Faktoren in der behandelten und der unbehandelten Gruppe gleich verteilt sind.

Das eben konstruierte Beispiel ist eine *randomisierte Kontrollstudie*. Allgemeiner gesprochen werden hierbei die zu behandelnden Subjekte zufällig aus einer Gesamtheit von Probanden ausgewählt, wobei der Behandlungsstatus den einzelnen Subjekten im Idealfall verborgen bleibt, zum Beispiel durch den Einsatz eines Placebo-Präparats in der Kontrollgruppe. In der Medizin hat sich seit den 1940er Jahren diese Form des „Experiments“ als Goldstandard für die Etablierung der Wirkung von Medikamenten und Therapiemethoden etabliert. Der Wirksamkeitsnachweis per randomisierter Kontrollstudie ist nunmehr seit über 30 Jahren Voraussetzung für die Zulassung von Medikamenten in Deutschland wie andernorts. Das Unvermögen, einen solchen Nachweis zu liefern, hat beispielsweise dazu geführt, dass in Frankreich homöopathischen Arzneimitteln dieses Jahr die Kassenerstattung gestrichen wurde.

Die empirische Sozialforschung ignorierte die Methodik der randomisierten Kontrollstudie lange Zeit. Üblich waren hier sogenannte Beobachtungsstudien, in der Medizin als epidemiologische Studien bekannt. Methodisch gesehen unterscheiden diese sich von den randomisierten Kontrollstudien dadurch, dass der Behandlungsstatus den an der Studie beteiligten Subjekten nicht zufällig zugewiesen wird. Als ein prominentes Beispiel dieser empirischen Methodik in der ökonomischen Entwicklungsländerforschung mag ein viel zitierter Aufsatz von Christopher Udry über landwirtschaftliche Produktivität in Burkina Faso aus dem Jahr 1996, veröffentlicht im *Journal of Political Economy*, dienen. Hier wird die Hypothese untersucht, dass landwirtschaftliche Haushalte ihre produktiven Ressourcen nicht optimal einsetzen im Hinblick darauf, dass innerhalb eines Haushalts Ackerstücke, die im Eigentum der Frau sind, weniger intensiv bewirtschaftet werden als solche, die im Eigentum des Mannes sind. Udry vergleicht hierzu den Einsatz von Saatgut, Arbeit und Düngemitteln auf Ackerstücken von Männern und Frauen, die dem gleichen Haushalt angehören. Die Einflüsse von beobachtbaren Unterschieden, beispielsweise der chemischen Bodenzusammensetzung verschiedener Flurstücke, werden hierbei mit statistischen Methoden einbezogen (oder „kontrolliert“). Im Gegensatz zu einer randomisierten Kontrollstudie ist hier die Zuweisung von Ackerstücken auf Frauen und Männer jedoch alles andere als zufällig. Vielmehr ist sie das Resultat komplexer Erbgänge und vielfältiger bewusster Entscheidungen verschiedener Akteure, beispielsweise der Eltern der Eheleute. Insofern lässt sich mit den verfügbaren Beobachtungsdaten nicht ausschließen, dass es von Udry unbeobachtete Unterschiede in der Produktivität von Feldstücken im Eigentum von Frauen und Männern gibt, gemäß derer es effizient ist, die „weiblichen“ Äcker weniger intensiv zu bewirtschaften als die „männlichen“. Dies wäre zum Beispiel der Fall, wenn „weibliche“ Ackerstücke bei gleichem Saatgut-, Düngemittel- und Arbeitseinsatz trotz identischer chemischer Bodenzusammensetzung durchschnittlich weniger ertragreich wären als die „männlichen“.

Im Bereich der empirischen Ökonomik entwickelte sich spätestens seit den 1980er Jahren eine Debatte über die Glaubhaftigkeit empirischer Forschungsergebnisse, die auf Beobachtungsstudien basieren. So klagte der renommierte Ökonometriker und Handelsökonom Edward Leamer im Jahr 1983 in dem Aufsatz *Let's Take the Con Out of Econometrics* (deutsch etwa „Lasst uns die Ökonometrie vom Schwindel befreien“), dass kaum ein Ökonom die Datenanalysen seiner Kollegen ernst nehme, da die meisten Ergebnisse verschwänden oder sich sogar umkehrten, wenn man auch nur kleine Details der jeweiligen statistischen Analysemethodik ändere.

Während Leamer ein verstärktes Augenmerk auf sogenannte Sensitivitätsanalysen forderte, ging eine damals junge Generation von amerikanischen Arbeitsmarktökonominnen das Kausalitätsproblem mit innovativen empirischen Forschungsdesigns an, die Kausalität dadurch etablieren, dass institutionelle Zusammenhänge ausgenutzt werden, die in einer zufälligen Zuweisung des „Behandlungsstatus“ resultieren. Ein bekanntes Beispiel ist die Arbeit des Chicago-Ökonomen Robert LaLonde aus dem Jahr 1986, der ein Lotterieverfahren zur Zuweisung von Plätzen in einem vom Staat finanzierten beruflichen Weiterbildungsprogramm ausnützt, um dessen kausalen Effekt auf spätere Beschäftigung und Verdienst zu bestimmen. Die Lotterie kam zum Einsatz, da die öffentlichen Mittel für das Weiterbildungsprogramm nicht ausreichten, um allen Bewerbern, die die Zugangsvoraussetzungen erfüllten, einen Platz bereitzustellen. In ähnlicher Weise schätzte der MIT-Ökonom Joshua Angrist 1990 den kausalen Effekt von Militärdienst auf den späteren beruflichen Werdegang junger amerikanischer Männer mit Hilfe des Lotteriemechanismus, mit dem US-Soldaten für den Vietnamkrieg bestimmt wurden.

Es ist offensichtlich, dass die Anzahl interessanter Forschungsfragen, die sich mit der Methodik solcher *natürlichen Zufallsexperimente* angehen lassen, äußerst begrenzt ist, insbesondere im Bereich der Entwicklungsländerforschung. Ein erwähnenswertes Beispiel in diesem Zusammenhang ist jedoch der Aufsatz der Preisträgerin Esther Duflo (gemeinsam mit Raghavendra Chattopadhyay) aus dem Jahr 2004, in dem die Autoren zeigen, dass weibliche Dorfvorsteherinnen in Indien häufiger in lokale Infrastruktur investieren, auf die Frauen größeren Wert legen als Männer, zum Beispiel Toiletten. Hierbei nutzen Sie die gesetzliche Quotenregelung aus, gemäß derer in einem Drittel der Dörfer des Bundesstaats West Bengal nur weibliche Kandidaten zur Wahl zugelassen sind. Die Regierung implementiert diese Quotenregelung so, dass in einem gegebenen Wahljahr jedes dritte Dorf von einer alphabetisch geordneten Liste ausgewählt wird - ein Verfahren, das die Autoren glaubhaft als „so gut wie zufällig“ bezeichnen. Die methodische Innovation dieser Arbeit liegt beim Forschungsdesign, das eine eigens konzipierte Datenerhebung von geschlechtsspezifischen Präferenzen sowie die Bereitstellung von verschiedenen lokalen öffentlichen Gütern in zahlreichen Dörfern beinhaltet und diese Primärdaten mit der Regierungsliste der für Frauen reservierten Dörfer kombiniert. Die vorhergehenden, in den USA angesiedelten Arbeitsmarktstudien basieren hingegen ausschließlich auf Sekundärdaten, an deren Erhebung die Wissenschaftler nicht selbst beteiligt waren.

Der methodische Unterschied zwischen solchen *natürlichen Zufallsexperimenten* und einer *randomisierten Kontrollstudie*, die man ebenso kontrolliertes Zufallsexperiment nennen könnte, ist, dass bei der Kontrollstudie der auswertende Forscher nicht nur die Datenerhebung gestaltet, sondern auch auf die Intervention und insbesondere den Zufallsprozess, gemäß dem der Behandlungsstatus zugewiesen wird, die

„Randomisierung“, Einfluss nimmt, während er beim natürlichen Zufallsexperiment rückwirkend Zufallselemente ausnützt, die ein bestimmter institutioneller Zusammenhang generiert. Es ist offensichtlich, dass sich mit randomisierten Kontrollstudien prinzipiell eine Vielzahl interessanter Forschungsfragen angehen lassen – vorausgesetzt dem Forscher gelingt es, an der Implementierung relevanter Maßnahmen hinreichend beteiligt zu werden.

Zusammenfassend lassen sich die gängigsten Forschungsdesigns in den empirischen Wirtschaftswissenschaften hierarchisch wie folgt typologisieren: bei der randomisierten Kontrollstudie steuert der Wissenschaftler die Randomisierung, beim natürlichen Zufallsexperiment erfolgt sie hingegen von anderer Stelle. Bei der Beobachtungsstudie liegt die Zuweisung des Behandlungsstatus ebenfalls nicht in der Hand des Forschers und ist darüber hinaus nicht zufällig. Eine Position zwischen den beiden letztgenannten Designs nimmt das *natürliche Experiment* ein, wo die Zuweisung des Behandlungsstatus nicht zufällig ist aber sich im Zuweisungsprozess Elemente finden lassen, die den Behandlungsstatus als „so gut wie zufällig“ erscheinen lassen. Eine beliebte Anwendung sind hier beispielsweise die ökonomischen Konsequenzen von Regenfallschwankungen, die auf die landwirtschaftliche Produktion in Entwicklungsländern einen großen Einfluss haben. Der aufmerksame Leser wird an dieser Stelle beanstanden, dass die eben diskutierte Studie von Duflo nach dieser Klassifizierung streng genommen „nur“ ein natürliches Experiment und kein natürliches Zufallsexperiment ist – allerdings wird man zustimmen, dass bei Chattopadhyay und Duflo die Vergleichbarkeit von „behandelten“ und Kontroll-Dörfern eher gegeben ist als z. B. von Dörfern, die im selben Jahr guten bzw. mangelhaften Niederschlag erfahren haben. Oder vielleicht doch nicht? Die eben aufgestellte Behauptung soll verdeutlichen, dass grundsätzlich jedes empirische Forschungsdesign, das keine strenge Randomisierung beinhaltet, hinsichtlich der Vergleichbarkeit von behandelter und Kontrollgruppe und somit auch im Hinblick auf die Bestimmung eines kausalen Effektes einer Intervention grundsätzlich angreifbar ist.

Im Rahmen der oben erwähnten Debatte über die Glaubhaftigkeit empirischer Forschungsergebnisse sind im Hinblick auf den Ansatz der Nobelpreisträger zwei Beiträge von 1994 und 1995 in der Zeitschrift *World Bank Research Observer* bemerkenswert: Hier brechen der Berkeley Ökonom Paul Gertler (mit John Newman und Laura Rawlings) und Michael Kremer eine Lanze für randomisierte Kontrollstudien, wobei Gertler auf Regierungsprogramme zur Schulbildung und Familienplanung der 70er, 80er und frühen 90er Jahre hinweist, die bereits randomisierte Evaluationen beinhalten, während Kremer argumentiert, dass Ökonomen selbst auf die Gestaltung von Pilot-Programmen Einfluss nehmen und sie durch randomisierte Zuteilung evaluieren sollten. Hierbei zeigt er, dass dies im Bereich der Bildungsforschung in Entwicklungsländern schon vereinzelt durchgeführt wurde – allerdings nicht von Ökonomen.

Kurz darauf waren es diese beiden Wissenschaftler, die als erste Ökonomen randomisierte Kontrollstudien in Entwicklungsländern durchführten. Gertler (Berkeley) gemeinsam mit Jere Behrman und Petra Todd von der University of Pennsylvania und dem Yale Ökonomen T. Paul Schultz konzipierten 1997 eine randomisierte Einführung und Evaluierung des Transferzahlungs- und Kinderförderungsprogramms *Progresá* der mexikanischen Regierung. Michael Kremer hingegen darf als Vater des *randomisierten Feldexperiments* in der Entwicklungsökonomie gelten. So werden randomisierte Kontrollstudien genannt, die

im wirtschaftlichen Alltag der beteiligten Subjekte angesiedelt sind und wo der Wissenschaftler auch an der Gestaltung der Intervention selbst beteiligt ist. Was die Implementierung betrifft sind sie üblicherweise von kleinerem Umfang als landesweite Regierungsprogramme wie Progresca. Typologisch sind sie abzugrenzen von ökonomischen *Laborexperimenten*, wo das Verhalten menschlicher Subjekte unter synthetischen Laborbedingungen untersucht wird. Diese wurden bereits in den 1960er Jahren, unter anderem vom bisher einzigen deutschen Wirtschafts-Nobelpreisträger Reinhard Selten, entwickelt.

Kremer und seine damaligen Doktoranden forcierten den experimentellen Forschungsansatz auf breiter Front. Die Erfolgsaussichten waren durchaus ungewiss. Aus heutiger Sicht ist es unbegreiflich, dass nicht wenigen seiner ersten experimentellen Studien, die er ab 1995 schwerpunktmäßig im ländlichen Kenia durchführte, die Veröffentlichung in einschlägigen Fachzeitschriften versagt blieb. Darunter befindet sich auch ein Feldexperiment mit Abhijit Banerjee, das den Effekt eines zusätzlichen Lehrers in indischen Dorfschulen auf Unterrichtsbesuch und Lernerfolg bestimmt, oder ein Aufsatz von Kremer (mit seiner Schülerin Christel Vermeersch), in dem bedeutende Effekte einer Schulspeisung auf Vorschulbesuch in Kenia nachgewiesen werden. Drei weitere Studien aus dieser Zeit erschienen erst mit sieben bzw. zehn Jahren Verzögerung. Ein Konferenzbeitrag Kremers über Feldexperimente im Schulbildungsbereich von Entwicklungsländern aus dem Jahr 2003 liest sich stellenweise wie ein Verteidigungsplädoyer der experimentellen Methodik gegenüber den damals noch dominierenden Beobachtungsstudien und natürlichen Experimenten. Dies verdeutlicht eindrucksvoll, wie schwer es wissenschaftlichen Fachcommunities – auch in den Wirtschaftswissenschaften – mitunter fällt, neue Paradigmen zuzulassen und zu honorieren, und in welchem Ausmaß auch der Wissenschaftsbetrieb von Moden und einer Tendenz zum Festhalten am *status quo* geprägt ist.

Die experimentelle Revolution in der Entwicklungsökonomie

Die ersten in der ökonomischen Fachgemeinschaft Aufsehen erregenden Publikationen, die auf randomisierten Evaluierungen in Entwicklungsländern basieren, erschienen im Jahr 2004, ein methodisch geprägter Aufsatz von Kremer (mit seinem Schüler Edward Miguel) über ein Entwurmungs-Programm in Kenia in der Zeitschrift *Econometrica*, sowie Gertlers Arbeit über Progrescas Wohlfahrtseffekte im *American Economic Review*. Unter anderem aufgrund der darin wiedergegebenen empirischen Evidenz fand Progresca in der Folge Nachahmung in mehr als 30 Ländern weltweit.

Es folgte ein regelrechter Boom von randomisierten Feldexperimenten in der ökonomischen Entwicklungsländerforschung. Grob kategorisieren lässt sich die mittlerweile in die Tausende gehende Vielzahl von Arbeiten mit randomisierten Forschungsdesigns hinsichtlich des Ausmaßes der Intervention und des Theoriebezugs der empirischen Forschung.

Wie schon ausgeführt waren bereits bei den allerersten randomisierten Evaluationen von Entwicklungsökonominnen einerseits im Maßstab überschaubare Feldexperimente und andererseits ein landesweites Regierungs-Sozialprogramm vertreten. In den Feldexperimenten wird häufig vergleichend getestet, welche Maßnahme in einem spezifischen Kontext am besten geeignet ist, um ein bestimmtes Entwicklungsziel zu

erreichen. Bei Kremers ersten Studien in Kenia waren die Ziele Schulbesuch und Lernerfolg, die Maßnahmen, die nacheinander erprobt wurden, Schulbücher, Flipcharts, Leistungsanreize für Lehrer und schließlich eine Entwurmungstherapie für Schüler. Nur die letzte dieser Maßnahmen war kosteneffizient und verbesserte nicht nur die Gesundheit der Kinder spürbar, sondern auch deren Schulbesuch und Lernerfolg. Es war die Zusammenarbeit mit einer finanziell gut ausgestatteten niederländischen Nichtregierungsorganisation, die es Kremer ermöglichte, all diese Maßnahmen durchzutesten. Esther Duflo bezeichnete dieses Aufspüren und Nachweisen von wirksamen Maßnahmen, das eher einen beharrlichen Charakter und Gespür für den jeweiligen Kontext als Brillanz auf dem Gebiet der mathematischen Modellbildung erfordert, als „Klempnerei“ in der ökonomischen Armutsforschung. Bei Regierungsprogrammen ist der Forscher an der Art der Maßnahme üblicherweise kaum beteiligt und die Evaluierung steht im Vordergrund. Typischerweise werden hierbei nicht alternative Maßnahmen im Vergleich getestet, sondern im Mittelpunkt steht die Wirksamkeit und Kosteneffizienz eines spezifischen großangelegten Programms. Ein aktiver junger Forscher auf diesem Gebiet ist Kremer-Schüler Karthik Muralidharan (San Diego), der beispielsweise große Verbesserungen sowohl für die Staatskasse, also auch für Leistungsempfänger, durch die Einführung von e-governance Maßnahmen in Indien nachgewiesen hat. Teilweise gelingt es Forschern, auch innerhalb von landesweiten Regierungsprogrammen experimentelle Variation einzuführen. So liegen eine Reihe von Studien unter Beteiligung Abhijit Banerjees über Indonesien vor, wo mehrere zentrale Elemente eines staatlichen Lebensmittel-Grundsicherungsprogramms unter Einbeziehung von Harvard- und MIT-Wissenschaftlern variiert wurden, sei es das Verfahren zur Bestimmung von Leistungsempfängern oder die Einbeziehung privater Dienstleister bei der Auslieferung der Nahrungsmittel vor Ort.

Eine weitere Differenzierung experimenteller Arbeiten lässt sich hinsichtlich des Theoriebezugs vornehmen. Während bei der „Klempnerei“ die Überprüfung theoretischer Hypothesen der Wirtschaftstheorie zugunsten praktischer Aspekte der Armutssenkung in den Hintergrund tritt, wurden in den letzten 15 Jahren zahlreiche raffinierte Forschungsdesigns implementiert, die die Bestimmung ansonsten schwierig zu messender Effekte ermöglichen, die in der ökonomischen Theoriebildung eine bedeutende Rolle spielen. Ein prominentes Beispiel, veröffentlicht 2009, ist das Feldexperiment von Duflo und Banerjees Schüler Dean Karlan (mit Jonathan Zinman) über adverse Selektion und moralisches Risiko (*moral hazard*) in Kreditmärkten, wo zunächst angebotene und letztlich im Vertrag abgeschlossene Kreditzinssätze unabhängig voneinander unter Kunden eines südafrikanischen Finanzdienstleisters randomisiert werden. In der Folge erhoben die Wissenschaftler Rückzahlungsdaten. Dieses Forschungsdesign erlaubte es erstmals, in überzeugender Weise zentrale Elemente der äußerst einflussreichen theoretischen Hypothesen zu asymmetrisch verteilter Information in Kreditmärkten zu testen, die 1981 von Nobelpreisträger Joseph Stiglitz veröffentlicht worden waren. Allerdings bringen solche Arbeiten eher akademischen Ruhm als einen Nutzen für die Armutsbekämpfung. Eine Stellung zwischen diesen Extremen nehmen Feldexperimente ein, die durch eine für das Leben der Betroffenen relevante Intervention Verhaltensparameter bestimmen, die der Theoriebildung entspringen und für den wirtschaftlichen Erfolg des Einzelnen eine wichtige Rolle spielen. Häufig werden hier Hypothesen aus der Verhaltensökonomik (*behavioral economics*) überprüft, die Abweichungen vom Lehrbuchmodell des rationalen

homo oeconomicus zum Gegenstand haben. Diese verhaltensökonomischen Elemente sind ein wichtiger Teil der Forschungsagenda von Banerjee und Duflo, die in ihrem bekannten Buch *Poor Economics* argumentieren, dass Mangel an Information, kognitive Überlastung und pessimistische Erwartungen wichtige Ursachen von chronischer Armut sind. So argumentieren Duflo und Kremer (mit Jonathan Robinson) in einem Aufsatz von 2011, dass das Entscheidungskalkül kenianischer Kleinbauern eine Verzerrung zugunsten der Gegenwart (*present bias*) aufweist, da ihre im Normalfall viel zu niedrige Nachfrage nach produktivitätssteigernden Düngemitteln insbesondere durch zeitlich begrenzte Rabatte während der Nach-Erntezeit, nicht aber während der Anbauperiode, erheblich gesteigert werden kann.

Grundsätzlich neu an Feldexperimenten ist die Arbeitsweise des Entwicklungsforschers. Bei Beobachtungsstudien wurde hauptsächlich mit nicht selbst erhobenen Daten (sog. Sekundärdaten) gearbeitet. In einer Minderzahl von Fällen war der Forscher an der Erhebung selbst beteiligt, was üblicherweise ein gewisses Maß an Arbeit im Feld beinhaltet. Andererseits gab es – eher vereinzelt – Politikberatung, wo Wissenschaftler meist eher indirekt Einfluss auf die Gestaltung einzelner Entwicklungsprojekte nehmen konnten. Beim Feldexperiment hingegen ist der Forscher sowohl bei der Gestaltung der Intervention als auch ihrer Implementierung und im Nachgang unmittelbar beteiligt. Außerdem führt er statistische Erhebungen in ihrem Umfeld durch. Dies führt zu mehr Teamarbeit, Kollaborationen mit Entwicklungsorganisationen und insgesamt einer größeren Verwurzelung in entwicklungspraktischer Arbeit im Feld als der von Schreibtischarbeit geprägte frühere Forschungsstil; und oft zu direkten Verbesserungen im Leben der von Armut Betroffenen. Um diese Arbeitsweise zu professionalisieren und Skaleneffekte auszunutzen, gründeten Banerjee und Duflo gemeinsam mit einem weiteren Kollegen 2003 am MIT das *Poverty Action Lab*, ein weltweit aktives Forschungszentrum mit mittlerweile 400 Angehörigen, das die Planung und Durchführung von RCTs und wissenschaftlichen Befragungen unterstützt. Die gemeinnützige Organisation wird heute durch namhafte Geber gefördert wie die Stiftungen von Google, der Familien Gates, Hewlett und MacArthur, sowie durch die US-amerikanische und die britische Regierung.

Das Ausmaß von RCTs in der entwicklungsökonomischen Forschungslandschaft, das zumindest teilweise als Messgröße für den Einfluss der von den Nobelpreisträgern vorwärtsgetriebenen Methodik auf die Forschungsdisziplin gelten kann, lässt sich auch bibliographisch belegen. Während im Jahr 2000 in den fünf renommiertesten volkswirtschaftlichen Fachzeitschriften 10 Prozent der veröffentlichten Artikel aus der Entwicklungsökonomie kamen und keiner davon eine randomisierte Kontrollstudie beinhaltete, waren es 2015 zwölf Prozent, wovon jeder dritte einen RCT beinhaltet. Andererseits weist der Weltbankökonom David McKenzie darauf hin, dass in den drei führenden entwicklungsökonomischen Fachzeitschriften nur jeder achte Aufsatz einen RCT beinhaltet. Dies zeigt, dass die Methodik der Nobelpreisträger ein hohes Gewicht insbesondere in der entwicklungsökonomischen Forschungsspitze erlangt hat, während sie im Forschungsoutput, den die Disziplin insgesamt produziert, deutlich weniger vertreten ist. Unstrittig ist hingegen, dass der experimentelle Ansatz die Denkkategorie des „idealen“ – will sagen randomisierten – Experiments in der empirischen Entwicklungsforschung als Goldstandard etabliert hat und die Mehrzahl nicht-experimenteller empirischer Arbeiten darauf Bezug nimmt.

Jüngere Entwicklungen in der experimentellen Entwicklungsforschung

Die drei Nobelpreisträger haben sich nicht nur um die Einführung randomisierter Feldexperimente im Entwicklungskontext verdient gemacht, sondern waren in den letzten 20 Jahren auch stets an vorderster Front bei ihrer methodischen Weiterentwicklung. Im Vordergrund stand hierbei stets das Bemühen um die Bestimmung der tatsächlichen Effekte verschiedenster Entwicklungsmaßnahmen, einschließlich ihrer Verbreitung in der Wissenschaftsgemeinschaft und Öffentlichkeit. Erhebliche methodische Fortschritte wurden insbesondere auf folgenden Problemfeldern erzielt:

- *Spillovers*: in ihrem Aufsatz über ein Entwurmungs-Programm unter kenianischen Schulkindern zeigen Kremer und sein Schüler Miguel, dass vorhergehende Studien aus dem Bereich der medizinischen Forschung die Effekte solcher Programme deutlich unterschätzen, da nicht behandelte Kontrollsubjekte von der Behandlung ihrer Schulkameraden erheblich profitieren. Die Autoren zeigen, wie solche Spillover-Effekte mit statistischen Verfahren geschätzt werden können und wie intelligente Randomisierungsdesigns dieser Herausforderung begegnen können.
- Gestaltung von Datenerhebungen: hier wurden Standards insbesondere mittels sogenannter Güteberechnungen etabliert. Forscher sind demnach angehalten, die Größe einer Zufallsstichprobe, zum Beispiel die Anzahl von Haushalten, die nach einer Intervention befragt werden, nicht willkürlich, sondern von der statistischen Verteilung der Variablen von Interesse und den zu überprüfenden Hypothesen abhängig zu wählen.
- Publikationsverzerrung (*publication bias*): eine oft geäußerte Sorge betreffend den Publikationsprozess des Wissenschaftsbetriebs ist, dass nur signifikante Ergebnisse zur Veröffentlichung gelangen, während so genannte Null-Ergebnisse in den Schubladen und Papierkörben der Wissenschaftler verschwinden. Wenn beispielsweise von 50 Studien zur Forschungsfrage „kann Zugang zu modernen Finanzdienstleistungen Armut lindern?“ nur fünf einen signifikant positiven Effekt finden und veröffentlicht werden, die anderen 45 aber unbemerkt bleiben, gibt die Gesamtheit der veröffentlichten Studien einen äußerst verzerrten Eindruck von der Armutswirkung eines verbesserten Zugangs zu Finanzdienstleistungen. In der Medizin wurde dieses Problem schon vor längerer Zeit erkannt und deshalb Register eingeführt, wo Studiendesigns schon vor der Durchführung registriert werden müssen, um später in renommierten Journalen veröffentlichen zu können. Während dies nicht verhindert, dass Null-Ergebnisse nicht aufgeschrieben werden oder aus anderen Gründen nicht zur Veröffentlichung kommen, so bieten diese Register immerhin die Möglichkeit, die Gesamtheit aller Studien zu einer bestimmten Fragestellung zu bestimmen. Nachdem das US-amerikanische *National Institute of Health* bereits im Jahr 2000 ein Internet-basiertes Archiv für randomisierte Kontrollstudien (RCTs) in der Medizin startete, folgte die Fachvereinigung *American Economic Association* im Jahr 2012 mit einer ähnlichen Plattform für RCTs in den Sozialwissenschaften. Esther Duflo ist eine Direktorin dieses Archivs. Seit letztem Jahr ist die Registrierung in diesem Archiv verpflichtend für alle Arbeiten, die RCTs beinhalten und in den Zeitschriften dieser Fachvereinigung zur Publikation eingereicht werden. Als weitere Maßnahme, maßgeblich von Schülern der Preisträger initiiert, hat die führende Fachzeitschrift *Journal of Development Economics* 2018 ein *registered reports* Verfahren eingeführt, bei dem Studien bereits vor Generierung der Ergebnisse zur späteren Veröffentlichung

eingereicht werden können. Der Begutachtungsprozess, bei dem üblicherweise über 90 Prozent der eingereichten Aufsätze abgelehnt werden, findet also unabhängig von den Ergebnissen statt. Bildhaft gesprochen werden hier die Gutachter der Zeitschrift wie einst Odysseus an den Mast der Tugend gebunden, die Qualität einer Studie unabhängig von ihren Ergebnissen zu bewerten.

- *Data mining* oder *cherry picking* auch *fishing* genannt: schon Edward Leamer warf seinen Kollegen vor, dass in viele ökonomische Analysen ohne offensichtliche Gründe Annahmen einfließen, die auf die Ergebnisse erheblichen Einfluss haben. Es besteht der Verdacht, dass der Wissenschaftler schlussendlich dasjenige statistische Modell oder diejenige Ergebnisvariable wählt, die die für das Fachpublikum interessantesten Ergebnisse liefert – ohne die anderen Ergebnisse, die im Forschungsprozess generiert wurden, zu erwähnen. Diese „Gefahr“ besteht auch bei randomisierten Kontrollstudien, zum Beispiel bei der Definition der Ergebnisvariable oder dem statistischen Auswertungsverfahren. Unter anderem deshalb sind im Bereich der Medikamentenzulassung in den USA bereits seit über 20 Jahren Standards zu einem vorgelagerten Analyseplan (*pre-analysis plan*) verpflichtend. Hierbei muss der Wissenschaftler schon vor Durchführung der Studie alle Details der Intervention, Randomisierung, Datenerhebung und auch der statistischen Auswertung in einem Register hinterlegen. In der abschließenden Veröffentlichung muss nachgewiesen werden, dass diesem Plan in allen Aspekten gefolgt wurde. Salopp gesprochen lässt sich der Wissenschaftler dadurch wie einst Odysseus vorab an den Mast der Tugend binden, um unlauteren Versuchungen bei der späteren Datenanalyse zu entgehen. Auch das RCT Register der *American Economic Association* beinhaltet vorgelagerte Analysepläne.
- Verallgemeinerbarkeit: ein verbreiteter Vorbehalt gegenüber dem randomisierten Feldexperiment ist, dass sein Ergebnis zwar für einen spezifischen Kontext zutreffend und glaubhaft ist – man spricht von hoher interner Validität – jedoch wenig Aufschluss über die Wirksamkeit einer Maßnahme in einem möglicherweise nur geringfügig anders gestalteten Kontext gibt – mangelnde externe Validität. Wenn ausreichend große Studienzahlen vorliegen, können erstens sogenannte Meta-Analysen kontextübergreifende Einsichten liefern. Systematische Meta-Analysen sind in den letzten 15 Jahren in der Folge methodisch deutlich solider und häufiger geworden. Durch die Register können auch nicht publizierte Resultate besser erfasst werden. In diesem Zusammenhang ist auch die 2008 gegründete *International Initiative for Impact Evaluation* (3IE) zu nennen, die, finanziert von der Weltbank, systematisch Überprüfungs- und Metastudien fördert und in Auftrag gibt. Zweitens wird immer häufiger dieselbe Forschungsfrage in verschiedenen Kontexten erkundet, z.B. die Auswirkungen von Zugang zu Kredit, und die Ergebnisse systematisch verglichen. Drittens werden bestimmte Interventionen mittlerweile gleichzeitig in verschiedenen Ländern von koordinierten Teams durchgeführt und die Ergebnisse gemeinsam veröffentlicht, so zum Beispiel ein Programm zur Förderung der Ärmsten durch kostenlose Überlassung von Vieh, berufliches und lebenspraktisches Training, sowie zeitlich begrenzte Transferzahlungen.

Kritiker des experimentellen Forschungsansatzes

Zwei namhafte Kritiker des experimentellen Forschungsansatzes sind der Wirtschafts-Nobelpreisträger von 2015, Angus Deaton, und der frühere Weltbank-Ökonom Martin

Ravallion, der einflussreiche Arbeiten auf dem Gebiet der Armutsmessung vorgelegt hat. Ein Hauptkritikpunkt betrifft die Forschungsfragen, die mit randomisierten Studien angegangen werden. Ravallion beklagt in diesem Zusammenhang, dass eine bedingungslose Präferenz für RCTs über andere Methoden die Menge der Forschungsgegenstände unangemessen beschränke und dass wichtige Fragen der wirtschaftlichen Entwicklung auf der Strecke blieben. Er argumentiert, dass randomisierte Designs vor allem für Interventionen, die private Güter betreffen, geeignet sind, weniger hingegen für öffentliche Güter und Infrastruktur. Dies verleite Wissenschaftler ebenso wie Entwicklungsorganisationen wie die Weltbank, den Blick von mit RCTs nicht erforschbaren Fragen abzuwenden. Interessant ist in diesem Zusammenhang eine Stellungnahme der deutschen Gesellschaft für Internationale Zusammenarbeit (GIZ) von 2012, sie werde experimentelle Forschung nicht zur Richtschnur für die eigene Arbeit machen, denn sie eigne sich nur für einen schmalen Anwendungsbereich.

Ein weiterer Kritikpunkt betrifft die externe Validität. Die meisten RCTs werden in Zusammenarbeit mit wenigen Nicht-Regierungsorganisationen (NGOs) durchgeführt, die typischerweise eine straffe Organisation aufweisen und forschungsaffin geprägt sind, in der Praxis sind dort häufig promovierte Absolventen renommierter Universitäten anzutreffen. Diese seien jedoch weder für das Gros der NGOs in der Entwicklungsszene noch für staatliche Akteure repräsentativ. Das heißt wir erfahren von vielen RCTs Effekte, die möglicherweise wenig aussagekräftig dafür sind, was in der Breite der Entwicklungspraxis bei einer gleichgearteten Intervention eintreten würde.

Schließlich gibt es ethische Vorbehalte, insbesondere, dass einem Großteil der Studien-Beteiligten ein Programm vorenthalten wird, entweder dauerhaft oder temporär. Schon bei Progresa wurde die Öffentlichkeit erst nach Abschluss der eineinhalbjährigen Evaluationsphase im Jahr 2000, als das Programm universell wurde, über die Randomisierung informiert – harsche Kritik aus der Presse und der mexikanischen Öffentlichkeit war die Folge.

Schließlich verschlingen randomisierte Evaluierungen erhebliche Ressourcen, die für Entwicklungsmaßnahmen selbst eingesetzt werden könnten. In der Forschungsszene gewinnt man darüber hinaus zeitweise den Eindruck, dass es zu einem regelrechten Wettrennen zwischen den Forscherteams gekommen ist, wer sich die größere Erhebung und noch größer angelegte Intervention leisten kann.

Werdegänge der Preisträger

Michael Kremer wurde 1964 in New York geboren. Er schloss 1985 einen Bachelor in Sozialwissenschaften an der Harvard University ab. Von jeher motiviert, die Lebenssituation in armen Ländern zu verbessern, arbeitete er im Anschluss ein Jahr als Volontär an einer Schule im Westen Kenias. Im Anschluss leitete er drei Jahre die Geschäfte der von ihm gegründeten gemeinnützigen Organisation *World Teach*, die Volontariate als Lehrer in Entwicklungsländern vermittelt. Dort blieb er bis 2017 Aufsichtsratsvorsitzender. Er kehrte an die Harvard University zurück, um ein Doktorandenstudium in Ökonomie aufzunehmen, das er 1992 abschloss. Während dieser Zeit und im Anschluss als Assistenzprofessor am benachbarten *Massachusetts Institute of Technology* (MIT) beschäftigte er sich als Theoretiker mit sehr allgemeinen historischen und makroökonomischen Aspekten von Entwicklung, beispielsweise wie die Größe von

zusammenhängenden Landmassen auf dem Globus im Laufe der Menschheitsgeschichte unterschiedliche Entwicklungsniveaus und Bevölkerungsdichten in den verschiedenen Erdteilen bedingt haben. Zunächst eher als Hobby und seinem persönlichen Interesse für die Schulsituation in Entwicklungsländern geschuldet begann er ab 1995, im Anschluss an eine Urlaubsreise nach Kenia, in der Gegend, wo er einst als Lehrer volontiert hatte, als Partner einer christlichen NGO zunächst sehr kleine, dann aber auch größere randomisierte Feldexperimente an Schulen durchzuführen. In den USA publizierte er während dieser Zeit weiterhin sehr erfolgreich zu makroökonomischen Aspekten der Entwicklung, was ihm 1999 die Rückkehr nach Harvard als Professor einbrachte. Dort hat er seit 2003 die Gates Professur für Entwicklungsländerforschung inne.

Neben seinen Pionierleistungen auf dem Gebiet der Feldexperimente war er federführend an der Entwicklung des *Advance Market Commitment*-Mechanismus der globalen Impf-Initiative GAVI beteiligt. Hierbei verpflichten sich öffentliche Geber, koordiniert von der Weltbank und UNICEF, Pharmafirmen dreistellige Millionenbeträge für die Entwicklung und Verbreitung von Impfstoffen zu einem niedrigen Festpreis in Entwicklungsländern zu zahlen. In diesem Rahmen haben mehrere Länder des globalen Nordens beispielsweise 1.5 Milliarden Dollar bereitgestellt, um die Entwicklung und Verbreitung von Impfstoffen gegen Pneumokokken-Infektionen, einer Hauptursache von Kindersterblichkeit, voranzutreiben. Im Jahr 2010 gründete er *Development Innovation Ventures* innerhalb der US-Entwicklungsbehörde USAID. Dieser Fonds verfügt über ein jährliches Budget von 23 Millionen US Dollar und vergibt Mittel für innovative Entwicklungsprojekte in einem für jedermann offenen Wettbewerbsverfahren.

Esther Duflo, geboren 1972 in Paris, war sechs Jahre alt, als sie in einem Comic Mutter Theresa und ihre karitative Arbeit in Calcutta sah. Ihr Interesse an Armut und Entwicklungsländern wurde durch ihre Mutter geweckt, die als Kinderärztin bei mehreren humanitären Einsätzen mitwirkte. Abhijit Banerjee, geboren 1961, wusste im Alter von sechs hingegen genau, wie das Leben der Armen aussah. Sein Vater war Ökonomieprofessor an einem College in Calcutta. Dort erwarb Banerjee einen Bachelorabschluss. Es folgte ein Master von der Jawaharlal Neru Universität in Delhi und 1988 ein Doktorat von Harvard, wo er beim Ökonomie-Nobelpreisträger Eric Maskin studierte. Nach Stellen an Princeton und Harvard wechselte er 1993 ans MIT, wo er seit 2003 die *Ford Foundation International* Professur innehat. Er machte sich zunächst einen Namen als Theoretiker und behandelte vielfältige informationsökonomische Themen, nur teilweise mit Entwicklungsländerbezug.

Duflo studierte zunächst Geschichte und später, inspiriert von einer wissenschaftlichen Hilfstätigkeit beim Harvard-Ökonomen Jeffrey Sachs während eines Praktikums in Moskau, Volkswirtschaftslehre in Paris, wo sie 1995 einen Masterabschluss erwarb. Sie ging in der Folge ans MIT, wo sie 1999 das Doktorat erwarb. Kremer und Banerjee waren dort ihre Lehrer, letzterer auch ein Betreuer ihrer Doktorarbeit. Sie blieb am MIT, wo sie seit 2005 die Abdul Latif Jameel Professur für Armutsbekämpfung und Entwicklung innehat. Ihre Doktorarbeit wie auch alle weiteren frühen Publikationen basieren durchweg auf der Methode des natürlichen Experiments. Ihre erste experimentelle Arbeit, die allerdings unter Angestellten einer US-Universität angesiedelt ist und die Nachfrage nach freiwilliger Altersvorsorge zum Gegenstand hat, erschien 2003. Ihr erster experimenteller Aufsatz mit Entwicklungsländerbezug, gemeinsam mit Banerjee verfasst, erschien 2007 –

und wieder geht es um verbesserten Schulunterricht, hier um zusätzliche Lehrer und Computer-unterstütztes Lernen in Indien. Es folgte eine Vielzahl einflussreicher Arbeiten zu verschiedensten Entwicklungsthemen, die Feldexperimente zum Gegenstand haben. Nach eigenen Angaben war Duflo bislang an mehr als 70 Feldexperimenten beteiligt. Sie ist von den drei Preisträgern die leidenschaftlichste und sichtbarste Advokatin des experimentellen Forschungsansatzes. Seit 2017 ist sie federführende Herausgeberin der Top-Zeitschrift *American Economic Review*. Banerjee und Duflo veröffentlichten 2011 das allgemeinverständlich geschriebene Buch *Poor Economics*, ein Bestseller, der eine Vielzahl ihrer eigenen und anderer experimenteller Arbeiten zusammenfasst. Banerjee und Duflo sind seit 2015 verheiratet und haben zwei Kinder.

Was kann der experimentelle Forschungsansatz zur globalen Armutsreduzierung beitragen?

Über unterschiedlichste Maßnahmen hinweg weisen zahlreiche randomisierte Feldexperimente Verbesserungen in einzelnen Lebensbereichen der Armen nach. Dagegen findet die überwältigende Mehrheit der Studien keine großen – oder wie Banerjee sagt „transformativen“ – Effekte auf Einkommen, Armut oder gar Wirtschaftswachstum. In diesem Zusammenhang äußerte Kremer in einem Interview sinngemäß: Langfristig gesehen mag Wirtschaftswachstum der wichtigste Faktor für die Beseitigung der Armut sein und vermutlich hat Deng Xiaping den größten Beitrag zur globalen Armutsreduzierung geleistet. Doch währenddessen sterben in jeder Minute Menschen, weil sie keinen Zugang zu sauberem Trinkwasser haben. Auch wenn Experimente lediglich helfen können, einzelne Missstände wie diesen zu lindern, sollten wir die Mühe nicht scheuen, sie durchzuführen.

Literatur

Banerjee, A. V., Duflo, E. (2011). *Poor Economics: a Radical Rethinking of the Way to Fight Global Poverty*. New York: PublicAffairs.

Banerjee, A. V., Duflo, E., Goldberg, N., Karlan, D., Osei, R., Parienté, W., Udry, C. (2015). "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science*, 348(6236): 1260799-1 - 1260799-16.

Chattopadhyay, R., E. Duflo (2004). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India." *Econometrica*, 72(5): 1409-1443.

Gertler, P. (2004). "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment." *American Economic Review*, 94 (2): 336-341.

Miguel, E., M. Kremer (2004). "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72 (1): 159-217.

LaLonde, R. J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data." *American Economic Review*, 76(4): 604–620.

Leamer, E. (1983). "Let's Take the Con Out of Econometrics." *American Economic Review*, 73(1): 31–43.