

**Yvette Bodry, Bernhard Boockmann, Philipp Kugler und Verena von Zitzewitz**

# **Experimentelle Public Policy Evaluation von Sozialpolitik**

## **AUF EINEN BLICK**

---

- Randomisierte Feldexperimente sind in der Sozialpolitikforschung eine besonders erfolversprechende Möglichkeit, um verlässliche Aussagen zu Ursache-Wirkung-Beziehungen zu treffen.
- In Deutschland ist die Anwendung randomisierter Feldexperimente in der Sozialpolitik bisher gering und auf bestimmte Bereiche beschränkt (insbesondere Bildungs- und Arbeitsmarktpolitik). Zur Evaluation anderer sozialpolitischer Maßnahmen werden meist Quasi-Experimente verwendet.
- In Frankreich wurde die Anwendung randomisierter Feldexperimente entscheidend durch die Politik vorangetrieben. Die starke Abhängigkeit von der staatlichen Finanzierung kann aber auch Nachteile haben.
- In den USA gibt es eine breite Anwendung randomisierter Feldexperimente in der Sozialpolitik. Bei der Finanzierung steht die Forschungsförderung an erster Stelle.
- Ausbaufähig sind in Deutschland die Finanzierung und Infrastruktur für die Durchführung randomisierter Feldexperimente, der Zugang zur Praxis und die Unterstützung durch die Politik.

## ABSTRACT

Ziel der Studie ist es, eine Übersicht über die experimentelle Public Policy Evaluation in der Sozialpolitik in Deutschland, in Frankreich und den USA seit dem Jahr 2011 zu vermitteln. Hierzu wurden Evaluationen und wissenschaftliche Untersuchungen recherchiert und die Rechercheergebnisse systematisch ausgewertet. Die Ergebnisse zeigen, dass randomisierte Feldexperimente in der Sozialpolitik in Deutschland im Vergleich zu den anderen Ländern nur vereinzelt vorgenommen werden. In vielen Bereichen der Sozialpolitik gibt es überhaupt keine experimentellen Feldstudien. Der Beitrag schließt mit Vorschlägen, wie die experimentelle Evaluation der Sozialpolitik verstärkt werden könnte.

## ABSTRACT

This study provides an overview of experimental public policy evaluation in social policy in Germany, France, and the US since 2011. For this purpose, a review of existing policy evaluations and scientific studies was conducted. The results show that randomised field experiments in social policy are carried out only sporadically in Germany compared to the two other countries. In many core areas of social policy there are no experimental field studies at all. The article concludes with suggestions on how experimental evaluation of social policies could be strengthened.

# Inhalt

1 ZIELSETZUNG UND VORGEHEN.....	1
1.1    ZIELSETZUNG.....	1
1.2    ABGRENZUNG DES UNTERSUCHUNGSGEGENSTANDES.....	2
1.3    METHODISCHES VORGEHEN.....	3
2 RANDOMISIERTE STUDIEN ZUR SOZIALPOLITIK IN DEUTSCHLAND.....	4
2.1    HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN.....	4
2.2    ERGEBNISSE DER ERFASSUNG.....	5
2.2.1    Art der Intervention und Evaluationszeiträume.....	5
2.2.2    Politikfelder.....	6
2.2.3    Wirkungsdimensionen.....	7
2.2.4    Durchführung der Experimente.....	7
2.2.5    Finanzierung der Studien.....	7
2.2.6    Zusammenfassung.....	8
3 QUASI-EXPERIMENTE ZUR SOZIALPOLITIK IN DEUTSCHLAND.....	8
3.1    RANDOMISIERTE STUDIEN VERSUS QUASI-EXPERIMENTE.....	8
3.2    ERGEBNISSE DER ERFASSUNG.....	9
3.2.1    Zeiträume.....	9
3.2.2    Politikfelder und Interventionen.....	9
3.2.3    Wirkungsdimensionen.....	10
3.2.4    Zusammenfassung.....	10
4 RANDOMISIERTE STUDIEN ZUR SOZIALPOLITIK IN FRANKREICH.....	10
4.1    HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN.....	10
4.2    ERGEBNISSE DER ERFASSUNG.....	11
4.2.1    Zeiträume.....	11
4.2.2    Politikfelder und Interventionen.....	12
4.2.3    Wirkungsdimensionen.....	13
4.2.4    Durchführende Institutionen.....	13
4.2.5    Finanzierung der Studien.....	13
4.2.6    Zusammenfassung.....	14
5 RANDOMISIERTE STUDIEN ZUR SOZIALPOLITIK IN DEN USA.....	14
5.1    HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN.....	14
5.2    ERGEBNISSE DER ERFASSUNG.....	15
5.2.1    Zeiträume.....	16
5.2.2    Politikfelder und Interventionen.....	16
5.2.3    Wirkungsdimensionen.....	17
5.2.4    Durchführung der Experimente.....	17
5.2.5    Finanzierung der Studien.....	17
5.2.6    Zusammenfassung.....	18
6 FORSCHUNGSDESIDERATE.....	18
LITERATUR.....	20

# 1 Zielsetzung und Vorgehen

## 1.1 ZIELSETZUNG

Wissenschaftliche Erkenntnisse über Wirkungszusammenhänge können dabei helfen, Sozialpolitik im Hinblick auf unterschiedliche Interventionsansätze und Zielgruppen besser zu fokussieren. Noch mehr als bisher ist es erforderlich, Evidenz auf der Mikroebene der Begünstigten zu gewinnen, um die individuellen Wirkungen sozialpolitischer Maßnahmen und Leistungen zu erfassen. Dieser Blickwinkel ist nicht zuletzt für die Akzeptanz von Sozialpolitik von größter Wichtigkeit.

Die Voraussetzung dafür sind verlässliche Forschungsergebnisse zu Ursache-Wirkung-Beziehungen. Randomisierte Feldexperimente sind hierbei ein besonders erfolgversprechender Ansatz. Durch die zufällige Zuordnung einer Intervention zu einer Treatmentgruppe<sup>1</sup> und spiegelbildlich den Ausschluss einer Kontrollgruppe von der Intervention können kausale Effekte glaubhaft identifiziert werden, weil die Zuordnung zur Intervention nicht systematisch mit Eigenschaften der einbezogenen Personen oder anderen Rahmenbedingungen zusammenhängt.

Im Bereich der Entwicklungspolitik hat der stärkere Fokus auf die Mikroebene verbunden mit der Nutzung randomisierter Feldexperimente die wissenschaftliche und politische Diskussion stark beeinflusst. Den Anstoß bildeten Publikationen wie „Poor Economics“ der späteren Nobelpreisträger\*innen Abhijit Banerjee und Esther Duflo (Banerjee/Duflo 2011) und die Einrichtung des Abdul Latif Jameel Poverty Action Lab (J-PAL) am Massachusetts Institute of Technology (MIT).

Randomisierte Feldexperimente könnten für die Sozialpolitikforschung ähnlich einflussreich werden. Von vornherein klar ist dabei, dass randomisierte Feldexperimente nicht für alle Fragen der Sozialpolitikforschung geeignet sind. Sie eignen sich am besten für genau definierte, abgrenzbare Interventionen auf der Mikroebene. Dagegen sind Maßnahmen mit substanziellen Rückwirkungen auf die gesellschaftliche Makroebene für die Methodik weniger geeignet.

Aufgabe der vorliegenden Expertise ist es, eine Übersicht über Studien der experimentellen Public Policy Evaluation in der Sozialpolitik in Deutschland und in zwei Vergleichsländern einschließlich der relevanten Akteure seit dem 01.01.2011 zu erstellen.

Das Ziel der Expertise ist die Beschreibung des Forschungsstandes und die Identifikation von Forschungslücken. Der Fokus liegt auf den sozialpolitischen Maßnahmen und Leistungen, deren Wirkungen in den vorhandenen Studien untersucht wurden, den dabei verwendeten Methoden, den am Experiment beteiligten Akteuren und der Finanzierung dieser Forschung. Eine Meta-Ana-

---

<sup>1</sup> Unter dem Begriff Treatment werden im Folgenden alle Arten von Interventionen oder Maßnahmen verstanden, deren Effekt auf die von der Intervention Betroffenen untersucht werden soll.

---

lyse der Studienergebnisse zu den Wirkungen ist dagegen nicht Gegenstand der Expertise. Zur Identifikation von Forschungslücken und -potenzialen werden Vergleichsländer herangezogen, in denen randomisierte Studien in größerem Umfang durchgeführt werden als in Deutschland.

## 1.2 ABGRENZUNG DES UNTERSUCHUNGSGEGENSTANDES

Die Durchführung von experimentellen Evaluationen in der Sozialpolitikforschung ist kein klar umrissener wissenschaftlicher Forschungsstrang. Vielmehr zeigt die im Folgenden vorgenommene Bestandsaufnahme der Literatur, dass sich die Forschungsarbeiten unterschiedlichen Forschungsrichtungen zuordnen lassen. Hierzu gehört beispielsweise die Evaluation von Arbeitsmarktpolitik, die ein etabliertes Forschungsgebiet darstellt. Ebenso bilden die in der Bildungsforschung durchgeführten Feldexperimente im Bereich der schulischen Bildung ein abgrenzbares Forschungsgebiet.

Unter Sozialpolitik werden im Folgenden alle staatlichen Maßnahmen zur Absicherung von Lebensrisiken, zum sozialen Schutz bestimmter Personengruppen und zur Senkung von Ungleichheit verstanden. Die Betrachtung umfasst auch die staatliche Bildungspolitik. Der Fokus der Untersuchung liegt auf Wirkungsstudien, die die Effekte sozialpolitischer Gestaltungen und Interventionen abschätzen und insofern den Beitrag der Sozialpolitik zur Lösung gesellschaftlicher Probleme untersuchen.

Der Auswahl der Vergleichsländer liegen folgende Kriterien zugrunde:

- Anzahl der durchgeführten Experimente,
- Breite der Abdeckung sozialpolitischer Teilbereiche durch die Experimente,
- ein Land muss ein weiteres EU-Land, das zweite ein OECD-Land sein.

Unter den EU-Ländern wurde Frankreich als Vergleichsland ausgewählt. Dort wurden relativ viele randomisierte Studien durchgeführt, und das Anwendungsfeld in der Sozialpolitik ist breit und umfasst nicht nur den Bildungsbereich und die Arbeitsmarktpolitik, sondern z. B. auch Transferleistungen und das Gesundheitsverhalten. Außerhalb der EU sind die USA das interessanteste Vergleichsland. Dies ergibt sich aus der großen Anzahl an randomisierten Feldstudien und der dadurch gewonnenen Umsetzungserfahrung sowie der breiten Anwendung in einer Vielzahl von sozialpolitischen Politikbereichen.

Kern dieser Studie sind randomisierte Feldexperimente, das heißt die zufällige Vergabe eines Treatments im Rahmen des jeweiligen Forschungsvorhabens. In Deutschland werden im Bereich der Sozialpolitikforschung Wirkungsstudien eher mit quasi-experimentellen Forschungsdesigns durchgeführt. Das bedeutet, dass die Vergabe des Treatments nicht durch einen im Rahmen des Forschungsvorhabens eingesetzten Zufallsmechanismus gesteuert wird, sondern durch einen nicht durch das Forschungsvorhaben gesteuerten Prozess (ein „natürliches Experiment“). Dieser bewirkt im besten Fall dasselbe wie eine gesteuerte Zufallsauswahl, nämlich dass die Verteilung aller relevanten beobachtbaren und nicht beobachtbaren Eigenschaften bei den Personen, die das Treatment erhalten bzw. nicht erhalten, identisch ist.

Von diesen Ansätzen sind solche Forschungsstrategien abzugrenzen, bei denen sich die Treatment- und Kontrollgruppe unterscheiden, die Unterschiede aber durch ein statistisches Verfahren (z. B. ein Regressionsmodell oder das Matching-Verfahren) herausgerechnet werden können. Dieser Ansatz beruht auf der Annahme, dass die Verteilung aller relevanten Eigenschaften bei den Personen, die das Treatment erhalten bzw. nicht erhalten, bedingt auf eine Reihe von beobachtbaren Merkmalen identisch ist (Annahme der bedingten Unabhängigkeit). Diese Annahme stellt hohe Anforderungen an die Informationsbasis und ist deshalb oft kritisch zu sehen. Studien, die diesen nicht-experimentellen Ansatz verwenden, werden in die nachfolgende Betrachtung nicht aufgenommen.<sup>2</sup>

### 1.3 METHODISCHES VORGEHEN

Zur Suche nach Experimenten werden im Bereich der randomisierten Experimente einschlägige Online-Archive genutzt. So listet die Datenbank des Abdul Latif Jameel Poverty Action Lab (<https://www.povertyactionlab.org/evaluations>) über 1.100 experimentelle Evaluationen weltweit auf, davon allerdings nur acht für Deutschland.

Die größte öffentliche Datenbank zur Registrierung von Experimenten im Bereich der Wirtschaftswissenschaften ist das 2013 gegründete RCT Registry der American Economic Association (AEA) (<https://www.socialscisearch.org/site/data>, siehe auch Abrams et al. 2020). Sie bietet einerseits eine Plattform, auf welcher ein externes Publikum ein Experiment nachvollziehen kann, und bedingt andererseits, dass Experimente vor ihrer Ausführung registriert werden müssen. Dies verhindert eine nachträgliche Manipulation der Datenerfassung. Die Datenbasis des AEA Registry reicht bis ins Jahr 2013 zurück und enthält insgesamt über 3.100 Studien. Davon sind über 100 Deutschland zugeordnet, wobei die Mehrzahl der Studien nicht die Sozialpolitik betrifft und die Zuordnung teilweise über die Verfasser\*innen und nicht über das Land, in dem das Experiment durchgeführt wurde, vorgenommen wird. Die vorliegende Zusammenstellung geht von allen Studien aus, die (1) im AEA RCT Registry registriert wurden, (2) in Form eines Arbeitspapiers, eines Zeitschriftartikels oder eines öffentlich verfügbaren Projektberichts publiziert wurden, (3) zwischen 2011 und 2022 durchgeführt wurden und (4) sich entweder Themen der Gesundheit, Bildung und Soziales (JEL Code I) oder der Arbeitsmarkt- bzw. Bevölkerungsökonomik (JEL Code J) widmen. Aufgrund der Selbstregistrierung umfasst die AEA Registry nicht alle randomisierten Studien, die in einem Land durchgeführt wurden. Außerdem sind die Angaben oft unvollständig. Neben der Suche nach Experimenten werden deshalb auch gezielte Suchläufe nach wissenschaftlichen Studien vorgenommen. Bei der Suche nach Quasi-Experimenten wird ausschließlich auf diese Weise vorgegangen. Hierzu werden Literaturdatenbanken wie z. B. Econbiz (<https://www.econbiz.de>) und IDEAS (<https://ideas.repec.org>) genutzt, die mit systematischen Suchstrategien wie z. B. Forward/Backward-Zitationen recherchiert werden.

Im Untersuchungsteil für Deutschland und Frankreich wird eine vollständige Übersicht über die bisher durchgeführten randomisierten Feldexperimente im Bereich der Sozialpolitik gegeben. Für die USA wurden aufgrund der weitaus höheren Fallzahl nur solche Studien aufgenommen, die in der AEA Registry registriert sind. Von einer zusätzlichen Recherche in Literaturdatenbanken wurde abgesehen, da auch ohne diesen Schritt eine umfassende Datenbasis gewonnen werden konnte. Da die Anzahl der quasi-experimentellen Studien um ein Vielfaches höher ist als die der randomisierten Studien, werden die Ergebnisse zu den quasi-experimentellen Verfahren stärker zusammengefasst. Beispielsweise werden alle Änderungen im Bereich des Mindestlohnes oder im Bereich der Rentenversicherung, die im Rahmen von quasi-experimentellen Studien analysiert worden sind, zusammengefasst dargestellt.

Die Studien werden in strukturierten Übersichtstabellen erfasst. Die Erfassungseinheit ist das Experiment oder Quasi-Experiment, nicht primär die wissenschaftliche Publikation, in der das Experiment wissenschaftlich verwertet wird. Wissenschaftliche Publikationen werden dabei jedoch erfasst und den jeweiligen Experimenten zugeordnet. Die Ergebnisse der Aufnahme sind

---

2 In der Literatur sind auch andere Abgrenzungen verbreitet. So versteht Kubbe (2021) unter Quasi-Experimenten alle Studien ohne Randomisierung des Treatments, aber mit einer Kontrollgruppe. Dies bezieht auch Ansätze ein, die die Annahme der bedingten Unabhängigkeit voraussetzen. Im Unterschied dazu wird hier unter einem quasi-experimentellen Vorgehen ausschließlich die Nutzung natürlicher Experimente verstanden, bei denen diese Annahme nicht erforderlich ist. Jankowski und Tepe (2022) grenzen natürliche Experimente und Quasi-Experimente voneinander ab, je nachdem, ob durch einen außerhalb des Forschungsvorhabens stattfindenden Prozess tatsächlich eine Randomisierung erfolgt ist.

---

in den EXCEL-Anhängen zu diesem Bericht enthalten.<sup>3</sup> Die Listen enthalten Informationen zur betrachteten sozialpolitischen Intervention, zur Art und Durchführung des (Quasi-) Experiments und zur jeweiligen Studie (z. B. die jeweils betrachteten Ergebnisvariablen). Ferner werden öffentlich verfügbare Links und/oder digital verfügbare Berichte angegeben. Die Inhalte der Datenbasis werden in Bezug auf experimentelle Methoden, Politikfelder, Auftraggeber und Politikenebene in den folgenden Kapiteln systematisch ausgewertet.

## 2 Randomisierte Studien zur Sozialpolitik in Deutschland

### 2.1 HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN

Die Studienlage für Deutschland, die im vorliegenden Abschnitt dargestellt wird, zeigt, dass randomisierte Feldexperimente im Bereich der Sozialpolitik relativ selten vorgenommen werden. Dies steht im Gegensatz zur Häufigkeit, mit der Wirkungsstudien sozialpolitischer Interventionen durchgeführt werden. So bezieht die Meta-Analyse zur Arbeitsmarktpolitik von Card et al. (2018) für Deutschland 42 Evaluationsstudien ein, von denen keine einzige ein randomisiertes Studiendesign verwendet. Dagegen enthalten 29% der insgesamt 41 Studien für die nordeuropäischen Länder experimentell erzielte Ergebnisse.<sup>4</sup> Es stellt sich daher die Frage nach den Ursachen für die vergleichsweise geringe Nutzung randomisierter Kontrollgruppenvergleiche.

Die Anwendung von randomisierten Feldexperimenten wird in Deutschland durch die Rechtsordnung begrenzt. Als übergeordnete Norm ist hier das Verfassungsgebot der Gleichbehandlung nach Art. 3 Abs. 1 GG zu nennen. Im Einzelfall bestehen vielfach Rechtsansprüche auf die jeweilige Förderung oder Leistung, so dass ein zufälliger Ausschluss davon rechtswidrig wäre. Randomisierte Feldexperimente sind also nur dort möglich, wo die Rechtsordnung der Verwaltung faktische Entscheidungsspielräume belässt oder wo Leistungen freiwillig an Leistungsempfänger\*innen vergeben werden. Letzteres ist insbesondere dann der Fall, wenn es sich nicht um staatliche, sondern um wohltätige Leistungen aus dem privaten und zivilgesellschaftlichen Bereich handelt. Allerdings gibt es auch in weiten Bereichen der staatlichen Sozialpolitik erhebliche Entscheidungsspielräume bei der Zuweisung und Vergabe von Maßnahmen und Leistungen, beispielsweise in der Arbeitsmarktpolitik der Bundesagentur für Arbeit (Bockmann et al. 2014).

Neben der Rechtsordnung und politischen Rahmenbedingungen könnte auch eine kritische Einstellung der Bevölkerung zu randomisierten Feldexperimenten eine Ursache für die relativ geringe Nutzung dieser Methodik in Deutschland sein. Nach den Ergebnissen von Fischer et al. (2019) ist jedoch eine Mehrheit (75%) der untersuchten Zufallsstichprobe aus allen wahlberechtigten Deutschen positiv gegenüber Evaluierungen eingestellt. Dabei hat es keinen Einfluss auf die Zustimmung zu Evaluierungen, ob diese explizit als „Experiment“ bezeichnet werden. Die Beschreibung des wissenschaftlichen Ansatzes als „Experiment“ wirkt anscheinend nicht stigmatisierend. Neben der Akzeptanz von randomisierten Feldexperimenten in der Bevölkerung ist auch die Akzeptanz bei den jeweils Beteiligten, beispielsweise bei den Auftraggebern von Studien, den Trägern der Leistungen sowie nicht zuletzt in der Zielgruppe, wichtig.

---

<sup>3</sup> Der Online Anhang kann unter <https://difis.org/institut/publikationen/publikation/52> heruntergeladen werden

<sup>4</sup> Eigene Auswertungsergebnisse auf Basis des Datenanhangs von Card et al. (2018), erhältlich unter <https://academic.oup.com/jeea/article/16/3/894/4430618#supplementary-data>.

---

Die Evaluation von staatlichen Programmen, Maßnahmen und Förderungen wurde in den letzten Jahren stärker in der Agenda der Bundes- und Landesregierungen verankert. Dazu gehören die vom Staatssekretärsausschuss Bürokratieabbau 2013 beschlossene „Konzeption zur Evaluierung neuer Regelungsvorhaben“, wonach finanzrelevante Gesetze nach drei bis fünf Jahren evaluiert werden müssen,<sup>5</sup> und die im Jahr 2019 beschlossene „Fortentwicklung der Evaluierungskonzeption der Bundesregierung“ (Bundesregierung 2019). Vorgaben oder Empfehlungen zur Nutzung bestimmter Methoden, die wie die randomisierten Feldexperimente einen besonders hohen Evidenzstandard erfüllen, werden dabei allerdings nicht gemacht. Auch die „Arbeitshilfe zur Evaluierung von Regelungen der Bundesregierung“ (Statistisches Bundesamt 2022) erwähnt randomisierte Feldexperimente nicht, sondern erläutert kausale Untersuchungsdesigns am Beispiel eines quasi-experimentellen Vorgehens. Unter den Studien, die im Rahmen des Politiklabors „Wirksam regieren“ der Bundesregierung durchgeführt wurden, sind ebenfalls nur wenige randomisierte Studien vorhanden (Bundesregierung o. J.). Trotz einer ausgeprägten Evaluationskultur ist die politische Unterstützung für randomisierte Feldstudien in der Sozialpolitik in Deutschland nach wie vor relativ schwach.

## 2.2 ERGEBNISSE DER ERFASSUNG

### 2.2.1 ART DER INTERVENTION UND EVALUATIONSZEITRÄUME

Insgesamt wurden für Deutschland im Bereich der Sozialpolitik 22 Experimente recherchiert, die die in Abschnitt 1.2 genannten Kriterien erfüllen (Siehe Anhang 1). Die Ergebnisse wurden in 45 wissenschaftlichen Publikationen veröffentlicht. Zwölf der Experimente implementieren ein sogenanntes Informationstreatment. Dieses besteht z. B. in der Zusendung einer Broschüre, einer telefonischen oder persönlichen Kontaktaufnahme oder einer kurzen Präsentation. Sie sollen auf Maßnahmen wie z. B. Weiterbildungsangebote hinweisen und dadurch den Zugang in die Maßnahmen beeinflussen. Bei den übrigen zehn Experimenten wird dagegen die Teilnahme an einer Maßnahme oder Unterstützung direkt durch einen Zufallsmechanismus vergeben. Dies kann z. B. die Teilnahme an einem Mentoring, einer Therapie, einem Bildungsangebot oder einer Leistung für Familien sein. Wir bezeichnen randomisierte Informationstreatments im Folgenden als niederschwellige Experimente, während die direkte Randomisierung der Teilnahme als substanzielles Experiment bezeichnet wird.

Abbildung 1 zeigt die zeitliche Verteilung der recherchierten Experimente. Nicht einbezogen sind alle Experimente, deren Durchführung bereits vor 2011 endete – darunter fallen allerdings nur sehr wenige Studien wie z. B. Arntz und Thomsen (2011).<sup>6</sup> Nicht einbezogen sind ferner Experimente, die so spät starteten, dass derzeit noch keine publizierten Ergebnisse vorliegen. Insgesamt lässt sich über den Beobachtungszeitraum seit 2011 kein klarer Trend in der Anzahl durchgeführter randomisierter Experimente im Bereich der Sozialpolitik feststellen. Während die bis 2014 begonnenen Experimente in der Mehrheit niederschwellige Treatments vergaben, überwiegen danach die substanziellen Treatments. Angesichts der geringen Zahlen ist diese Tendenzaussage allerdings mit Vorsicht zu interpretieren.

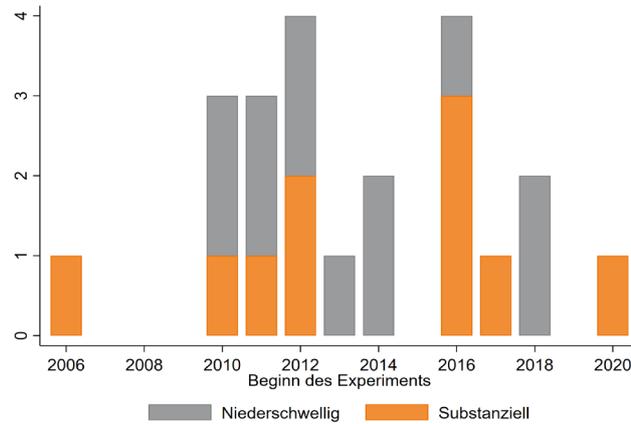
---

5 [https://www.bmwi.de/Redaktion/DE/Downloads/J-L/konzeption-zur-evaluierung-neuer-regelungsvorhaben-2013.pdf?\\_\\_blob=publicationFile&v=6](https://www.bmwi.de/Redaktion/DE/Downloads/J-L/konzeption-zur-evaluierung-neuer-regelungsvorhaben-2013.pdf?__blob=publicationFile&v=6)

6 <https://www.bundesregierung.de/breg-de/themen/wirksam-regieren>. Im Bereich der Sozialpolitik ist derzeit eine Evaluierung des Flexirentengesetzes in Arbeit, die aber offenbar auch kein experimentelles Forschungsdesign umsetzt.

---

Abbildung 1: Anzahl der Experimente nach Jahr des Beginns (Deutschland)

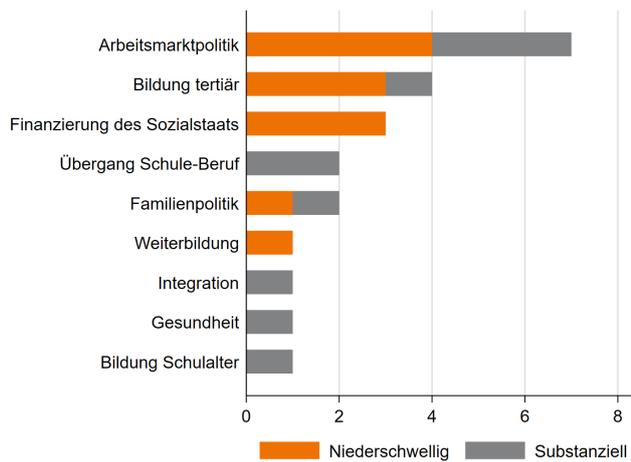


Quelle: eigene Darstellung

### 2.2.2 POLITIKFELDER

In der experimentellen Politik-Evaluation dominieren im Bereich der Sozialpolitik die Arbeitsmarkt- und Bildungspolitik (siehe Abbildung 2). Insgesamt sieben Studien entfallen auf den Bereich der schulischen oder beruflichen Erstausbildung, davon vier auf den tertiären Bildungsbereich, während für den frühkindlichen Bildungsbereich keine experimentellen Studien bekannt sind. Ebenfalls sieben Experimente wurden zur Evaluierung von Programmen und Maßnahmen im Bereich der Arbeitsmarktpolitik durchgeführt. Große Bereiche der Sozialpolitik wurden bisher nicht experimentell evaluiert. Dies betrifft insbesondere den Bereich der Sozialversicherung, der sozialen Grundsicherung und der sozialen Regulierung.

Abbildung 2: Anzahl der Experimente nach Politikbereichen (Deutschland)



Quelle: eigene Darstellung

Informationstreatments werden in Form von individuellen Anschreiben bei Fragen der Finanzierung des Sozialstaats genutzt sowie im tertiären Bildungsbereich, wobei es sich um Informationen zu bestimmten Studiengängen handelt. Auch bei den Leistungen der Arbeitsmarktpolitik und der Weiterbildung überwiegen die Informationstreatments.

### 2.2.3 WIRKUNGSDIMENSIONEN

Die Ergebnisvariablen der Studien sind heterogen und hängen in erster Linie davon ab, welche Ergebnisse durch die jeweilige Intervention im Beobachtungszeitraum beeinflusst werden können. Es finden sich darunter Persönlichkeitsmerkmale (Prosozialität, Altruismus, Risikoaversion), psychische Gesundheit und Gesundheitsverhalten, Teilnahme an bestimmten Bildungsgängen oder an beruflicher Weiterbildung, Bildungsergebnisse wie Noten oder Sprachkenntnisse und Zufriedenheitsmaße. Nur in fünf Fällen werden der Arbeitsmarktstatus oder das Einkommen betrachtet. Die betrachteten Ergebnisse hängen auch davon ab, über welchen Zeitraum die Teilnehmenden und Nichtteilnehmenden beobachtet werden. Beispielsweise wurden in den Publikationen zum Mentoring-Programm „Balu und Du“ sowohl kurzfristige Ergebnisse wie z. B. Selbsteinschätzungen, aber auch längerfristige Outcomes wie z. B. die Schulwahl mehrere Jahre nach der Teilnahme am Programm untersucht. Bei den Teilnehmenden am Experiment zum Selbstregulierungstraining Mental Contrasting with Implementation Intentions (MCII, deutsch „mentales Kontrastieren mit Durchführungsvorsätzen“) wurden sowohl kurzfristige Ergebnisse wie die Qualität der angefertigten Lebensläufe als auch die längerfristige Arbeitsmarktintegration betrachtet. Eine besonders langfristige Beobachtung und eine entsprechend differenzierte Betrachtung kurz- und langfristiger Ergebnisse war auch bei der Evaluation des Hausbesuchs-Programms „Pro Kind“ möglich.

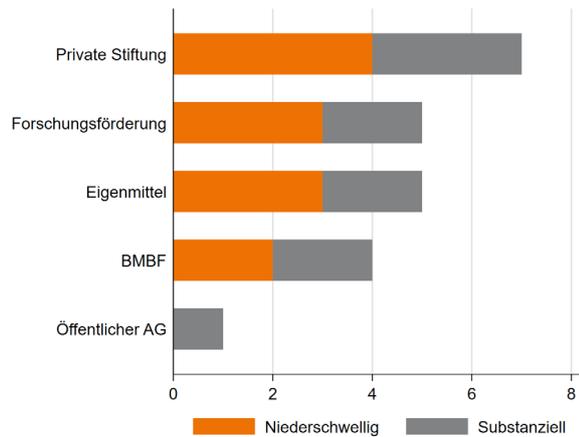
### 2.2.4 DURCHFÜHRUNG DER EXPERIMENTE

Die Durchführung und Evaluation der Experimente umfassen ein breites Spektrum forschender Einrichtungen. So waren bei der Durchführung von elf Projekten Universitäten verantwortlich oder beteiligt, darunter auch Wissenschaftler\*innen von ausländischen Universitäten wie der London School of Economics (LSE) und dem University College London (UCL). Das Institut für Arbeitsmarkt- und Berufsforschung (IAB) hat drei Experimente durchgeführt und bei zwei weiteren Experimenten mitgewirkt. Das ifo-Institut München und das Wissenschaftszentrum Berlin für Sozialforschung (WZB) waren bei jeweils vier Experimenten beteiligt, das Deutsche Institut für Wirtschaftsforschung (DIW) bei drei und das RWI sowie das briq-Institut jeweils bei zwei Experimenten. Zudem waren weitere wissenschaftliche Partner sowie Praxispartner, wie beispielsweise die Agentur für Arbeit, ein Landesbildungsministerium, eine gesetzliche Krankenkasse, Kirchen und andere Institutionen involviert.

### 2.2.5 FINANZIERUNG DER STUDIEN

Sieben von 22 recherchierten Experimenten wurden durch private Stiftungen finanziert oder kofinanziert (siehe Abbildung 3) Darüber hinaus tragen die Wissenschaftsförderung und die wissenschaftlichen Einrichtungen durch Eigenmittel erheblich zur Finanzierung von randomisierten Feldexperimenten bei. Die Deutsche Forschungsgemeinschaft (DFG) hat vier Projekte finanziert oder kofinanziert. Ebenfalls vier Projekte erhielten Fördermittel des Bundesministeriums für Bildung und Forschung (BMBF). Ferner haben einzelne Projekte Forschungsmittel des European Research Council (ERC), des britischen Economic and Social Research Council (ESRC) oder anderer Forschungsförderprogramme im Ausland erhalten. Nur eine Studie, die Evaluation des Programms Talentscouting NRW, kann als Auftragsforschung eines öffentlichen Auftraggebers eingeordnet werden. Studien mit niederschweligen und substanziellen Experimenten verteilen sich gleichmäßig auf die Finanzierungsquellen.

Abbildung 3: Anzahl der Experimente nach Art der Finanzierung (Deutschland)



Quelle: eigene Darstellung

### 2.2.6 ZUSAMMENFASSUNG

In Deutschland wurden bislang im internationalen Vergleich relativ wenige randomisierte Feldexperimente im Bereich der Sozialpolitik durchgeführt. Die Anwendung randomisierter Kontrollgruppenvergleiche ist in der Sozialpolitik in Deutschland derzeit auf bestimmte Bereiche beschränkt. Dazu gehören Bildungs- und Mentoring-Programme, die in der Regel durch private Stiftungen initiiert und finanziert werden. Mehrere randomisierte Studien finden sich auch im Bereich der Arbeitsmarktpolitik. Sie betreffen jedoch nicht unmittelbar das Regelinstrumentarium der Arbeitsmarktpolitik, sondern eher Umsetzungsvarianten oder Informationstreatments. Demgegenüber sind weite Bereiche der Sozialpolitik – insbesondere die Sozialversicherung, die staatliche Grundversicherung und die soziale Regulierung – bisher nicht durch randomisierte Feldexperimente erfasst.

## 3 Quasi-Experimente zur Sozialpolitik in Deutschland

### 3.1 RANDOMISIERTE STUDIEN VERSUS QUASI-EXPERIMENTE

Durch die Hinzunahme von Quasi-Experimenten zur Gruppe der betrachteten Studien vergrößert sich der Bereich der Interventionen, über deren Wirkung empirische Aussagen getroffen werden können. Neben spezifischen Interventionen in der Sozialpolitik, bei denen eine zufällige Vergabe vorstellbar ist, rücken so auch die Kernbereiche der sozialen Sicherung in den Blick. Quasi-experimentelle Verfahren basieren auf Beobachtungsdaten und nicht auf experimentellen Daten, die häufig spezifisch erhoben werden müssen. Daher können entweder Daten allgemeiner Haushalts- oder Personenbefragungen verwendet werden oder administrative Daten z. B. der Sozialversicherungsträger. Dadurch sind quasi-experimentelle Studien oft weniger aufwendig in der Durchführung als randomisierte Studien. Ferner bietet sich die Möglichkeit, teilweise weit zurückliegende Interventionen (bis hin zur Einführung der Sozialversicherung in Deutschland im 19. Jahrhundert, siehe Bauernschuster et al. 2020) zu betrachten und die Beobachtung über einen langen Zeitraum fortzusetzen. Die Verwendung bevölkerungsrepräsentativer Daten hat ferner den Vorteil, dass die Ergebnisse der Studien nicht wie bei den randomisierten Feldexperimenten von einem in der Regel zeitlich und lokal begrenzten Experiment auf andere Kontexte übertragen werden müssen.

## 3.2 ERGEBNISSE DER ERFASSUNG

### 3.2.1 ZEITRÄUME

Durch die Literatursuche wurden zwölf Gruppen von Quasi-Experimenten im Bereich der Sozialpolitik identifiziert (siehe Tabelle 1). Dabei umfasst jede Gruppe Quasi-Experimente eines Interventionsbereichs. Beispielsweise umfasst der Interventionsbereich „Änderungen beim Elternzeitananspruch“ sowohl Änderungen in der Anspruchsdauer in den 2000er Jahren als auch die Baby Jahr-Reform 1986 in der DDR. Für die zwölf Interventionsbereiche wurden 23 Beispielstudien in die Datenbank aufgenommen. Diese Studien sind zwischen 2012 und 2021 erschienen, der Zeitpunkt des Quasi-Experiments liegt jedoch in einigen Fällen weit davor. So untersucht eine Studie die Einführung der Pflichtkrankenversicherung für Arbeiter im Jahr 1877. Die meisten Studien setzen jedoch an neueren sozialstaatlichen Änderungen an, beispielsweise der Einführung des gesetzlichen Mindestlohns 2015, der Einführung des SGB II im Jahr 2005 oder der graduellen Einführung des G8 nach dem Jahr 2000.

**Tabelle 1: Gegenstände von Quasi-Experimenten in der Sozialpolitik**

Politikfeld	Interventionsbereich des Quasi-Experiments	Beispielstudien
Alterssicherung	Änderungen der Regeln zum Rentenzugang	Geyer/Welteke 2021
Krankenversicherung	Einführung einer verpflichtenden Krankenversicherung	Bauernschuster et al. 2020
	Beschränkter Zugang zur Krankenversicherung für Geflüchtete	Wenner et al. 2020
Arbeitslosenversicherung	Änderung der Regeln für den Eingliederungszuschuss (EGZ)	Boockmann et al. 2012
	Änderungen des Gründungszuschusses für Arbeitslose	Caliendo/Tübbicke 2021
Soziale Mindestsicherung	Organisationsformen im SGB II	Mergele/Weber 2020
Familienpolitik	Änderungen beim Elternzeitananspruch	Ziefle/Gangl 2014
Soziale Regulierung	Einführung des gesetzlichen Mindestlohns	Dustmann et al. 2020
	Änderungen beim Kündigungsschutz	Bauer et al. 2007
Berufsorientierung	Variation in der Verfügbarkeit von Berufsinformationszentren	Saniter et al. 2019
Tertiäre Bildung	Einführung von Studiengebühren	Bruckmeier/Wigger 2014

Quelle: Eigene Zusammenstellung

### 3.2.2 POLITIKFELDER UND INTERVENTIONEN

In einer Reihe von Studien werden grundlegende Gestaltungsoptionen im Bereich der Sozialversicherung betrachtet. So werden im Bereich der Rentenversicherung Zugangsregeln in den Rentenbezug für unterschiedliche Personengruppen untersucht. Ausgangspunkt sind dabei die Rentenreformen der letzten Jahrzehnte, z. B. durch das Rentenreformgesetz 1999 und das RV-Nachhaltigkeitsgesetz aus dem Jahr 2004. Dabei macht sich die Evaluation zunutze, dass die Reformen die Angehörigen benachbarter Geburtsjahrgänge unterschiedlich betrafen, die Abschaffung der Altersrente für Frauen beispielsweise erstmals den Jahrgang 1952, aber nicht den Jahrgang 1951. Im Bereich der Krankenversicherung ist es schwieriger, Vergleichsgruppen zu bilden. Genutzt wird der beschränkte Zugang zur gesetzlichen Krankenversicherung (GKV) für Geflüchtete und zwei politische Reformen (1997 und 2007), die die Wartezeit für den Zugang in die GKV verlängerten. Die Wirkungen der Leistungen der Arbeitslosenversicherung auf die Anreize zur (Wieder-) Aufnahme einer Beschäftigung wird anhand der Verkürzung der Bezugsdauer von Arbeitslosengeld I im Jahr 2006 analysiert.

Die Familienpolitik ist ein häufiges Anwendungsgebiet für Quasi-Experimente. Eine beispielhaft in den Überblick aufgenommene Studie betrachtet Verlängerungen des Elternzeitanpruchs zwischen 1986 und 1992 sowie Anpassungen 2001 und 2007 und untersucht die Effekte auf die Dauer von Erwerbsunterbrechungen. Grundlegende Umgestaltungen in der Bildungspolitik eröffnen ebenfalls die Möglichkeit zur Anwendung von Quasi-Experimenten. Eine Reihe von Studien nimmt beispielsweise die Verkürzung der Schulzeit im Gymnasium von neun auf acht Jahre nach 2000 zum Ausgangspunkt.

Weitere Studien beschäftigten sich mit der aktiven Arbeitsmarktpolitik der Bundesagentur für Arbeit bzw. der Jobcenter. Die Möglichkeit zu quasi-experimentellen Vorgehensweisen bieten Änderungen im Zugang zu arbeitsmarktpolitischen Maßnahmen, beispielsweise im Zuge der Hartz-Reformen in den frühen 2000er Jahren. Betroffen hiervon sind Maßnahmen wie z. B. die Eingliederungszuschüsse oder die Förderung von Selbständigen.

Schließlich lassen sich soziale Regulierungen quasi-experimentell evaluieren. Dies betrifft den Kündigungsschutz, wozu die Schwellenwertregelungen genutzt werden können, die in den 1990er und 2000er Jahren mehrfach geändert wurden. Es betrifft insbesondere den gesetzlichen Mindestlohn, zu dessen Wirkungen seit seiner Einführung im Jahr 2015 eine Vielzahl von quasi-experimentellen Studien erschienen ist.

### 3.2.3 WIRKUNGSDIMENSIONEN

Insgesamt 14 der 23 Studien – und damit ein weit höherer Anteil als bei den Studien auf Basis randomisierter Feldexperimente – betrachten den Erwerbstatus oder Entgelte als Ergebnisvariablen. Dies hängt mit der häufigen Verbreitung von administrativen oder Registerdaten als Datenquelle zusammen, in denen diese Ergebnisse enthalten sind. Weitere Ergebnisvariablen sind z. B. Bildungsabschlüsse, Gesundheitsausgaben und Lebenszufriedenheit.

### 3.2.4 ZUSAMMENFASSUNG

Im Anwendungsbereich quasi-experimenteller und randomisierter Studien auf unterschiedliche Teilbereiche der Sozialpolitik in Deutschland gibt es nur geringe Überschneidungen. Quasi-Experimente werden verwendet, um die Regelleistungen der Sozialversicherung und der staatlichen Grundsicherung sowie die sozialen Regulierungen zu evaluieren. Dabei werden die wichtigsten Bereiche der Sozialpolitik abgedeckt. Randomisierte Feldexperimente dienen dagegen zur Wirkungsanalyse einiger Förderprogramme, insbesondere solcher, die von privaten Stiftungen aufgelegt und finanziert werden. Nur sehr vereinzelt gibt es randomisierte Feldstudien in solchen Bereichen des Regelgeschäfts, in denen lokale Ermessensspielräume bestehen, wie z. B. bei der Anwendung von Eingliederungsvereinbarungen in der Arbeitslosenversicherung.

## 4 Randomisierte Studien zur Sozialpolitik in Frankreich

### 4.1 HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN

Nach der Etablierung der randomisierten Experimente in den USA und in der Entwicklungsökonomik begann sich die Methode in Frankreich im Bereich der Sozialpolitik ab Mitte der 2000er Jahre zu etablieren.<sup>7</sup> Die Evaluation öffentlicher Förderprogramme durch randomisierte Experimente wurde insbesondere durch den Politiker Martin Hirsch (zu diesem Zeitpunkt Haut Commissaire aux Solidarités actives und Haut Commissaire à la Jeunesse und damit Regierungsangehöriger) und die MIT-Professorin und Nobelpreisträgerin Esther Duflo vorangetrieben. Während sich Hirsch für die Finanzierung von Experimenten und insbesondere für die Methode der Randomisierung einsetzte, gründete Duflo 2008 das europäische Büro des Forschungszentrums Abdul Latif Jameel Poverty Action Lab (J-PAL), das sich zum

---

<sup>7</sup> Für einen weitergehenden Abriss der Geschichte der randomisierten Feldexperimente in Frankreich siehe insbesondere de Cler (2021, S. 7ff.).

---

Ziel setzt, weltweit mithilfe randomisierter Experimente zur Armutsbekämpfung beizutragen.

Ebenfalls im Jahr 2008 wurde der Fonds d'Expérimentation pour la Jeunesse (FEJ) gegründet. Der FEJ setzt sich zum Ziel, Programme zu finanzieren und zu evaluieren, die zum Schulerfolg, zur Chancengleichheit und zur sozialen wie beruflichen Integration der unter 25-jährigen Jugendlichen in Frankreich beitragen. Die Ergebnisse der Programmevaluationen tragen auf nationaler und lokaler Ebene zur Reflektion über die flächendeckende Einführung oder Erweiterung von Programmen bei (FEJ 2019). Zehn Jahre nach seiner Gründung hat der FEJ knapp 1.000 Projekte mit einem Gesamtbudget von rund 260 Millionen Euro umgesetzt (FEJ 2019). Während der Begriff „Expérimentation“ eine Bandbreite qualitativer und quantitativer Programmevaluationen umfasst, so ermöglicht der Fonds doch auch in entscheidender Weise die Finanzierung randomisierter Experimente.

Die Förderung experimenteller Projekte in der französischen Sozialpolitik beruht auf der engen Zusammenarbeit zwischen Politik, privaten Akteuren und der Wissenschaft. Eine Befragung der Projektträger des FEJ aus dem Jahr 2018 ergab unter anderem, dass die im Rahmen des FEJ durchgeführten Experimente die Vertrautheit der Träger mit experimentellen Methoden förderten und sie besser in die Lage versetzten, nützliche Erkenntnisse daraus zu ziehen (James und Bricet 2021). Zugleich entwickelte sich der FEJ in Richtung einer größeren methodischen Vielfalt weiter und die experimentelle Methode verlor die privilegierte Stellung, die sie in den Anfangsjahren des FEJ hatte. Diese Entwicklung kann aus den im Folgenden dargestellten Ergebnissen unmittelbar nachvollzogen werden. Für den Rückgang der Anwendung randomisierter Feldexperimente werden einerseits der Abschied von Hirsch aus der französischen Regierung und der Rückgang der Finanzierung für den FEJ verantwortlich gemacht (Jatteau 2016, S. 175, Fn. 744; S. 185). Zum anderen wird auf den begrenzten Wert der Ergebnisse randomisierter Feldexperimente für die Politikgestaltung verwiesen (de Cler 2021, S. 15).

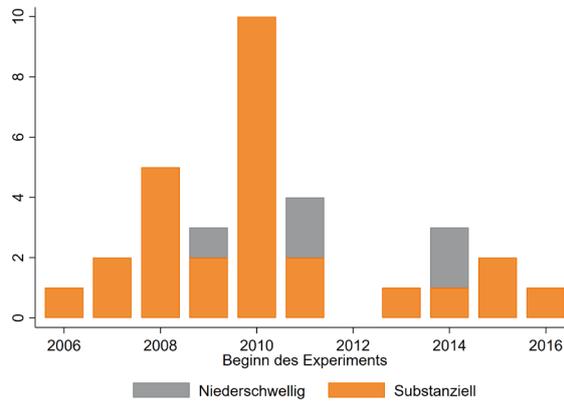
## 4.2 ERGEBNISSE DER ERFASSUNG

### 4.2.1 ZEITRÄUME

Für Frankreich wurden 32 randomisierte Experimente im Bereich der Sozialpolitik recherchiert, zu denen mindestens eine Publikation (Bericht, Artikel usw.) erschienen ist (siehe Anhang 3). Die Recherche ergab zudem sechs weitere Experimente, zu denen allerdings keine Veröffentlichungen auffindbar waren, teilweise weil die Studien erst 2021 begonnen wurden. Die Ergebnisse der 33 Experimente wurden in 52 Publikationen dokumentiert.

Die Blütezeit der randomisierten Experimente in der französischen Sozialpolitik fand Ende der 2000er und zu Beginn der 2010er Jahre statt (siehe Abbildung 4). Bis 2011 wurden 20 der insgesamt 33 recherchierten Experimente begonnen. Danach verloren randomisierte Studien in der französischen Sozialpolitik an Bedeutung. Seit 2014 wurden nur sechs der 32 recherchierten Experimente begonnen. Werden auch Projekte berücksichtigt, für die noch keine Projektberichte vorliegen, so lassen sich allerdings mindestens neun randomisierte Experimente der französischen Sozialpolitik in diesem Zeitraum verzeichnen und damit ähnlich viele wie in Deutschland. Nur fünf der 33 Experimente – und damit ein deutlich geringerer Anteil als in Deutschland – nutzen die Methode des Informationstreatments, während in den anderen Fällen die Teilnahme bzw. die Teilnahmeberechtigung am jeweiligen Treatment direkt randomisiert wird.

Abbildung 4: Anzahl der Experimente nach Jahr des Beginns (Frankreich)

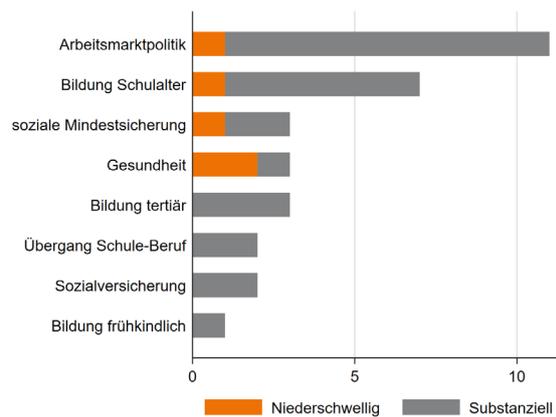


Quelle: eigene Zusammenfassung

#### 4.2.2 POLITIKFELDER UND INTERVENTIONEN

Ähnlich wie in Deutschland dominieren randomisierte Studien aus den Bereichen Bildungs- und Arbeitsmarktpolitik (13 bzw. 11 Experimente, siehe Abbildung 5). Innerhalb der Bildungspolitik betrifft die Mehrzahl der Experimente Interventionen im Bereich der allgemeinbildenden Schule. Bei der Arbeitsmarktpolitik liegt der Schwerpunkt auf Programmen für Ausbildungsabsolvent\*innen oder junge Arbeitslose, was vermutlich auf die Finanzierung durch den FEJ zurückzuführen ist. Anders als in Deutschland sind auch die soziale Mindestsicherung und die Sozialversicherung Gegenstand experimenteller Evaluierungen. Hierzu zählen Zuschüsse zur Krankenversicherung für Einkommensschwache (chèque santé), kostenlose gesundheitliche Beratungen, finanzielle Unterstützung und Mentoring für Jüngere, die benachteiligt oder in prekären Lebenslagen sind, sowie Beratungen zur Erhöhung der Inanspruchnahme von Sozialleistungen im Rahmen der „Caisses d’allocations familiales“ (Caf) oder der „revenu de solidarité active“ (RSA). Ferner wird eine Transferleistung untersucht, die beim Besuch einer Schwangerschaftsvorsorge gezahlt wird.

Abbildung 5: Anzahl der Experimente nach Politikbereichen (Frankreich)



Quelle: eigene Zusammenfassung

Die konkreten Interventionen sind hinsichtlich der Art und Dauer sehr unterschiedlich. Häufig handelt es sich um Beratungsleistungen oder ein Mentoring für Jugendliche oder ihre Eltern. Im Bildungsbereich geht es um spezifische Unterstützungen wie z. B. die Teilnahme an Sprachförderung, aber teilweise auch um die Teilnahme an einer längeren Beratung und Begleitung (z. B. „Groupement de Créateurs“) oder Bildungsgängen (z. B. Zugang zum Internat Sourdun). Daneben werden die Effekte finanzieller Leistungen (z. B. zum Erwerb

eines Führerscheines oder zur Unterstützung der Mobilität junger Arbeitsloser) untersucht.

### 4.2.3 WIRKUNGSDIMENSIONEN

Die Wirkungsdimensionen unterscheiden sich nach dem Interventionsbereich der Maßnahme. Die Studien zur Arbeitsmarktpolitik untersuchen in der großen Mehrheit die weiteren beruflichen Verläufe der Teilnehmenden, insbesondere den Eintritt in eine Beschäftigung oder Ausbildung nach Arbeitslosigkeit. Von den 52 berücksichtigten Studien betrachten 17 ausschließlich oder überwiegend diese „harten“ Indikatoren. Ferner wird der Bezug von Transferleistungen untersucht. Seltener werden „weiche“ Indikatoren, insbesondere solche, die einem Wechsel in Beschäftigung zeitlich vorlaufen, gewählt: Nur fünf Studien fokussieren auf Ergebnisvariablen wie z. B. subjektive Einstellungen, berufliche Aussichten, berufliche Wünsche oder das Suchverhalten am Arbeitsmarkt. In den Studien zu den Bildungsinterventionen wird ein relativ breites Spektrum von Ergebnisvariablen betrachtet. Dabei geht es um das Lern- und Sozialverhalten der Teilnehmenden, das Bildungsinteresse, die Elternbeteiligung, erlernte Kenntnisse z. B. in bestimmten Unterrichtsfächern, Schulleistungen und Schulerfolg sowie den Übergang in weiterführende Bildungsgänge wie z. B. ein Studium. Insgesamt 15 Studien betrachten diese Ergebnisvariablen.

Die Inanspruchnahme von Sozialleistungen ist in fünf weiteren Studien die betrachtete Wirkungsdimension. Vier Studien betrachten Ergebnisse aus mehreren Dimensionen gemeinsam, z. B. beruflicher Status, Wohnsituation, finanzielle Unabhängigkeit und materiel-ler Wohlstand. Nur sehr wenige Studien untersuchen gesundheitliche Outcomes.

### 4.2.4 DURCHFÜHRENDE INSTITUTIONEN

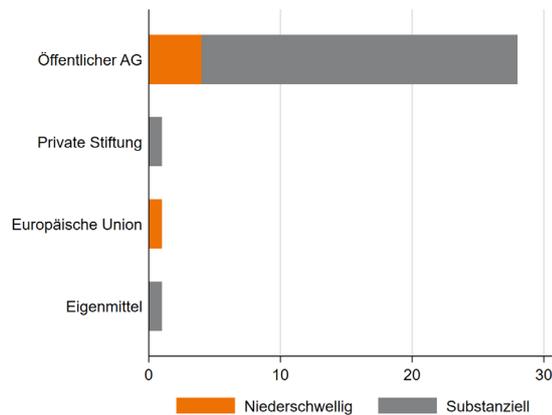
Als wichtige Akteure in der Programmevaluation treten das Forschungsinstitut des französischen Arbeitsministeriums Dares (Direction de l'animation de la recherche, des études et des statistiques), das ressortübergreifende Forschungszentrum DREES (Direction de la recherche, des études, de l'évaluation et des statistiques), das Forschungsinstitut Centre de recherche en économie et statistique (Crest), das Institut des politiques publiques (ipp), die Ecole d'économie de Paris, weitere Universitäten sowie das Forschungszentrum Abdul Latif Jameel Poverty Action Lab (J-PAL) auf. Hinsichtlich der involvierten Wissenschaftler\*innen fällt eine starke Konzentration auf die Wirtschaftswissenschaften und auf bestimmte Personen auf.

### 4.2.5 FINANZIERUNG DER STUDIEN

Die Finanzierung der Studien erfolgt fast ausschließlich durch öffentliche Auftraggeber (siehe Abbildung 6). Der FEJ trug maßgeblich zur Finanzierung von 20 der 33 randomisierten Experimenten bei. Dies betrifft insbesondere die bis 2011 begonnenen 20 Experimente, von denen 17 im Rahmen des FEJ stattfanden. Die durch den FEJ finanzierten Experimente zur Förderung von Jugendlichen reichen von Bildungsprojekten, Mentoringprojekten, über Projekte zur beruflichen Integration von Jugendlichen bis zur Förderung von Mobilität, Unterkunft oder der Inanspruchnahme von Gesundheitsleistungen.

Hinzu kommen Finanzierungen aus anderen Ministerien, wie dem Gesundheits- oder Wirtschaftsministerium, durch die Arbeitsagentur (Pôle Emploi) oder die Rentenversicherung. Diese Experimente analysieren beispielsweise die Leistungen von öffentlichen und privaten Jobcentern, die Nichtinanspruchnahme von Sozialleistungen oder Anreizprogramme in Bezug auf Vorsorgemaßnahmen. Private Stiftungen treten anders als in Deutschland bei der Initiierung, Durchführung und Finanzierung von Feldexperimenten praktisch nicht auf.

Abbildung 6: Anzahl der Experimente nach Art der Finanzierung (Frankreich)



Quelle: eigene Zusammenfassung

#### 4.2.6 ZUSAMMENFASSUNG

Die Durchführung randomisierter Experimente in der französischen Sozialpolitik begann um das Jahr 2010 und wurde gerade in diesem Zeitraum entscheidend durch die Finanzierung des Fonds d'Expérimentation pour la Jeunesse (FEJ) vorangetrieben. Der Schwerpunkt der Förderung durch den FEJ begünstigte dabei insbesondere die Umsetzung von sozialpolitischen Projekten für Jugendliche. Hinzu kommen weitere Projekte, beispielsweise aus der Arbeitsmarkt- oder Gesundheitspolitik. Die starke Unterstützung durch den Staat war jedoch nicht nachhaltig. Da andere Förderer wie z. B. private Stiftungen nicht einsprangen, ging die Anzahl der randomisierten Feldexperimente und der entsprechenden Publikationen wieder zurück.

## 5 Randomisierte Studien zur Sozialpolitik in den USA

### 5.1 HISTORIE UND INSTITUTIONELLE RAHMENBEDINGUNGEN

In den USA wurden randomisierte Feldstudien früher als in anderen Ländern durchgeführt. So untersuchten Thorndike und Woodworth schon im Jahr 1901 die Wirkungen mentaler Trainings in einer Reihe von Experimenten mit Schulkindern. In der Cambridge-Somerville Jugendstudie, dem ersten großen Feldexperiment in der Kriminologie, wurden zwischen 1942 und 1945 etwa 500 kriminalitätsgefährdete Jungen nach dem Zufallsprinzip einer Treatmentgruppe zugewiesen, die Beratungsdienste, medizinische Behandlung und schulische Nachhilfe erhalten konnte (Jamison 2019; Powers/Witmer 1951).

Vor allem die Große Depression befeuerte das Bedürfnis nach effektiven Methoden zur Evaluation von Sozialreformen, die den Bürgern der Vereinigten Staaten möglichst effektiv zugutekommen sollten (Oakley 1998). Es folgte das „Golden Age of Evaluation“ während der 1960er und 1980er Jahre. Randomisierte Feldexperimente fanden eine hohe öffentliche Akzeptanz (Jamison 2019). Eines der bekanntesten Programme, die aus dieser Zeit stammen, ist das Head Start Programm. In regelmäßigen Abständen wird bis heute in randomisierten Studien die Wirksamkeit der Maßnahmen getestet (Office of Head Start o.J.).

Die Beliebtheit randomisierter kontrollierter Studien zur Bewertung politischer Maßnahmen ging in den Vereinigten Staaten nach 1980 zurück (Oakley 1998). Auch in der Wissenschaft mangelte es an Enthusiasmus für randomisierte Experimente (Gueron/Rolston 2013).

Zur Trendwende der Wirtschaftswissenschaften in den 2000er zu datengestützten empirischen Studien trug bei, dass das U.S. Department of Health and Human Services und das Office of Management and Budget den Bundesstaaten 1992 mehr Freiheiten einräumten, gesetzliche Reformen zu erlassen (Gueron/Rolston 2013). Die Regierung unter Präsident Obama gründete das Social and Behavioral Sciences Team, welches 2014 begann in Zusammenarbeit mit Exekutivbehörden Experimente durchzuführen. Die Experimente basierten dabei größtenteils auf randomisierten kontrollierten Studien (Haskins 2017).

Darüber hinaus setzen sich Stiftungen für eine routinemäßige Evaluation von politischen Maßnahmen durch randomisierte kontrollierte Studien ein. Um dem Argument der hohen Kosten der Experimente entgegenzuwirken, gründete die Stiftung die „Coalition for Evidence-Based Policy’s Low-Cost RCT Competition“ und fördert zudem den Datenaustausch zwischen Forschenden, Behörden und nationalen Datenbanken (Buck/McGee, 2015). Auch die William T. Grant Foundation unterstützt Projekte, die Forschende und politische Entscheidungsträger und Entscheidungsträgerinnen in eine strukturierte und intensive Zusammenarbeit einbinden. Die Alfred P. Sloan Foundation leistet der Harvard University finanzielle Hilfe für die Entwicklung von neuen Methoden und Techniken, die die Robustheit und Aussagekraft randomisierter kontrollierter Studien erhöhen.

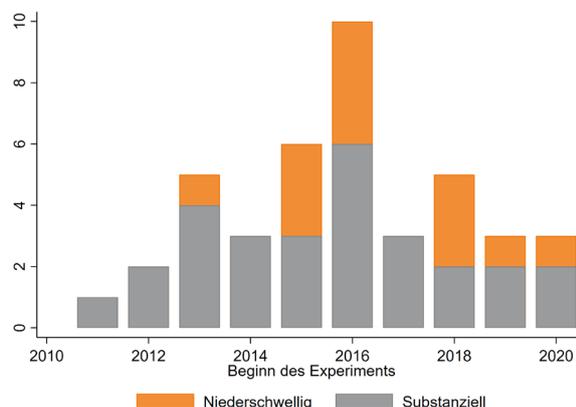
Während zu Beginn des ersten Jahrzehnts der 2000er Jahre die Methodik der randomisierten kontrollierten Experimente ein Wiederaufleben erfuhr, wuchs jedoch ebenso Misstrauen in die Evaluation dieser Experimente. Zu den Ursachen zählen auch heute noch Ergebnismanipulation bei der nachträglichen Auswertung der Versuchsdaten (p-Hacking) und die Selektion der Daten, welche der Öffentlichkeit zugänglich gemacht werden. Um dem entgegenzuwirken, wurde unter anderem das Register AEA Registry zur Vorabregistrierung von randomisierten Experimenten gegründet (Abrams et al. 2020).

## 5.2 ERGEBNISSE DER ERFASSUNG

### 5.2.1 ZEITRÄUME

Aufgrund der ausschließlichen Verwendung des AEA Registry für die Studienbasis aus den USA decken die hier erfassten Studien nur den Zeitraum ab 2013 ab. Insgesamt wurden zwischen 2013 und 2022 insgesamt 110 Experimente für die USA in der AEA RCT Registry registriert, die sich mit Interventionen im Bereich Gesundheit, Bildung und Soziales (JEL Code I) oder Arbeitsmarkt- bzw. Bevölkerungsökonomik (JEL Code J) befassen. Nach einer Sichtung der Projektbeschreibungen verblieben 78 Experimente mit sozialpolitischen Interventionen in der Datenbank. Von diesen liegen für 41 Experimente veröffentlichte Ergebnisse aus insgesamt 48 Publikationen vor.

Abbildung 7: Anzahl der Experimente nach Jahr des Beginns (USA)



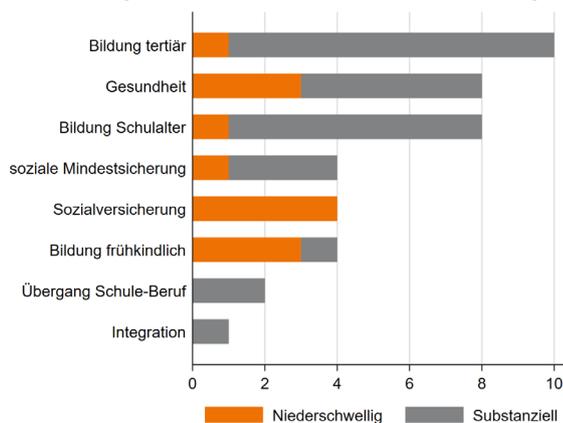
Quelle: eigene Zusammenfassung

Die experimentellen Studien in der Datenbasis wurden in den Jahren 2011 bis 2020 begonnen (siehe Abbildung 7). Der Rückgang der Zahlen nach 2016 kann damit zusammenhängen, dass für die in den letzten Jahren begonnenen Studien noch keine Ergebnisse publiziert sind. Insgesamt ist eine regelmäßige Durchführung von randomisierten Studien in den USA festzustellen. Die Laufzeit der Projekte ist sehr unterschiedlich und reicht von nur wenigen Monaten bis hin zu mehreren Jahren. In mehr als zwei Dritteln der Fälle (28 von 41 Experimenten) wird die Teilnahme an der Förderung und den Maßnahmen direkt randomisiert (substanzielles Treatment), in den übrigen Fällen werden niederschwellige Treatments vergeben.

### 5.2.2 POLITIKFELDER UND INTERVENTIONEN

Die Mehrzahl der Experimente betrifft den Bildungsbereich (siehe Abbildung 8). Maßnahmen im schulischen und tertiären Bereich stehen dabei mit zehn bzw. acht Experimenten an erster Stelle, aber mit vier Studien ist auch der frühkindliche Bildungsbereich vertreten. Acht Experimente sind dem Bereich der Gesundheit zuzuordnen. Auch zur sozialen Mindestsicherung und Sozialversicherung sind jeweils vier Experimente vorhanden. Im Gegensatz zu den beiden europäischen Ländern fehlt der Bereich der Arbeitsmarktpolitik bei den in der Datenbank enthaltenen Studien. Niederschwellige Treatments werden vor allem in sensiblen Bereichen wie Gesundheit, frühkindliche Bildung und Sozialversicherung verwendet. Möglicherweise stehen hier rechtliche Gründe einer direkten Randomisierung des Zugangs zu Leistungen und Maßnahmen entgegen.

**Abbildung 8: Anzahl der Experimente nach Politikbereichen (USA)**



Quelle: eigene Zusammenfassung

Auch innerhalb dieser Interventionsbereiche wird ein weites Spektrum von Maßnahmen, Förderungen und Vorgehensweisen abgedeckt. Im Bereich der Bildung geht es in einer Reihe von Experimenten um Interventionen im frühkindlichen und vorschulischen Bereich. Im schulischen Bereich finden sich z. B. Interventionen zur Schulwahl sowie Coaching- bzw. Tutoring-Programme. Im Hochschulbereich geht es unter anderem um die Wirkung von Online-Lerntools, Teilnahme an Coaching-Programmen und um die Wirkung von Rollenmodellen (also erfolgreichen Absolventinnen). Im Bereich der Gesundheit wurden einerseits eher niederschwellige Interventionen betrachtet, beispielsweise Online-Tools für den Zugang zu Medikamenten oder medizinischen Leistungen. Teilweise wurden aber auch medizinische Leistungen betrachtet, die über einen längeren Zeitraum durchgeführt wurden, beispielsweise ein Programm für Hausbesuche und medizinische Begleitung oder ein zweijähriges Wellness-Programm am Arbeitsplatz.

Die Experimente im Bereich der Sozialversicherung betrafen einerseits Maßnahmen zur Verbesserung der Entrichtung von Sozialbeiträgen, aber auch Informationen zur Prüfung der Renteneinkünfte. Vorhandene Studien beschäftigen sich ferner mit der Inanspruchnahme von Leistungen der sozialen Mindestsicherung sowie mit Transferzahlungen und Lebensmittelgutscheinen an

Haushalte mit geringem Einkommen. Auch Unterstützungen für die Zielgruppe der Wohnungslosen wurden experimentell evaluiert.

### 5.2.3 WIRKUNGSDIMENSIONEN

In den erfassten Experimenten wird eine Vielzahl unterschiedlicher Bildungsauscomes betrachtet. Im Bereich frühkindlicher Bildung geht es z. B. um kognitive und nicht-kognitive Testergebnisse und den Besuch von Kindergärten und Vorschulen. Im schulischen Bereich werden die Schülerleistungen (Kursnoten), Lese- Schreib- und mathematischen Fähigkeiten sowie Ergebnisse von Leistungstests betrachtet, aber auch das Leseverhalten. Vielfach wird der Einfluss der Interventionen an Bildungsübergängen gemessen, z. B. die Wahl von postsekundären Bildungsgängen unterschiedlicher Art und Qualität.

Bei den Gesundheitsoutcomes geht es um die Teilnahme an Präventionsleistungen und das Gesundheitsverhalten. Es werden ferner medizinische Diagnosen sowie (häufiger) die Inanspruchnahme von Gesundheitseinrichtungen und -leistungen (Arztbesuche, Besuch der Notaufnahme, stationäre Krankenhausaufnahmen, Dauer des Aufenthalts im Krankenhaus usw.) betrachtet. Eine weitere Gruppe von Outcomes sind finanzielle Ergebnisse, z. B. die Höhe der Inanspruchnahme öffentlicher Gesundheitsleistungen.

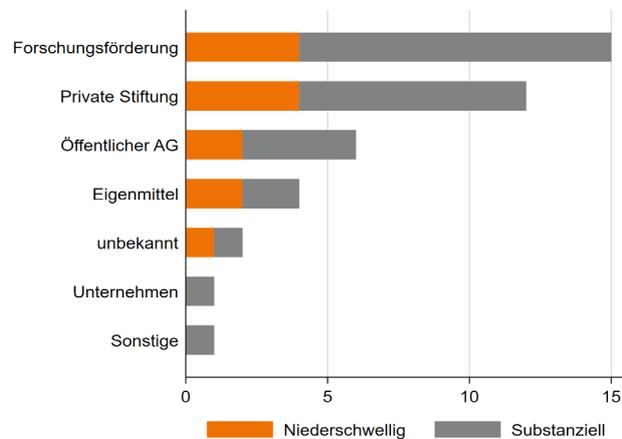
Arbeitsmarkt- und ökonomische Outcomes sind entsprechend den Politikfeldern selten. Nur in drei Studien werden Erwerbseinkommen oder Beschäftigung untersucht. Mehrere Studien untersuchen die Wirkung der jeweiligen Intervention auf kriminelles Verhalten, Straffälligkeit bzw. Rückfälligkeit und strafrechtliche Ergebnisse (z. B. Anzahl der Gefängnisaufenthalte).

### 5.2.4 DURCHFÜHRUNG DER EXPERIMENTE

Die starke Rolle der Wissenschaft bei der Durchführung der Experimente zeigt sich darin, dass die Ergebnisse fast durchweg in wissenschaftlichen Zeitschriften (darunter viele mit sehr hohem Impact Factor) publiziert wurden. An den Experimenten waren häufig außerdem projektspezifische Umsetzungspartner beteiligt. Zum Teil waren Forschungseinrichtungen im Bereich der experimentellen Forschung einbezogen (z. B. Wharton Behavioral Laboratory).

### 5.2.5 FINANZIERUNG DER STUDIEN

Die Projekte wurden durch eine Vielzahl an unterschiedlichen Akteuren finanziert, wobei die Forschungsförderung durch staatliche oder private Fördereinrichtungen die häufigste Finanzierungsquelle ist (siehe Abbildung 9). Häufig wurden die Projekte von verschiedenen Akteuren gleichzeitig finanziert. Dabei waren oft Ministerien und staatliche Institute sowie private Stiftungen beteiligt.

**Abbildung 9: Anzahl der Experimente nach Art der Finanzierung (USA)**

Quelle: eigene Zusammenfassung

### 5.2.6 ZUSAMMENFASSUNG

Die stetige Durchführung einer Vielzahl von randomisierten Feldstudien und die Heterogenität der beteiligten Akteure zeigt, dass eine große Anzahl an Akteuren Interesse daran hat, randomisierte Projekte durchzuführen und zu finanzieren. Der Fokus der Projekte liegt dabei hauptsächlich im Bildungs- und Gesundheitsbereich. Auch der Zugang zu sozialer Sicherung und Sozialversicherung ist Thema experimenteller Studien. Bei der Finanzierungsstruktur steht die Forschungsförderung an erster Stelle, wobei es häufig eine gemeinsame Finanzierung durch Forschungsförderung und staatliche oder nichtstaatliche Auftraggeber gibt. Bemerkenswert ist die Qualität des akademischen Outputs der Evaluationen.

## 6 Forschungsdesiderate

Randomisierte Wirkungsstudien spielen in der deutschen Sozialpolitikforschung bislang nur eine periphere Rolle. Dass das nicht so sein muss, zeigt die Erfahrung anderer Länder. In den USA und in Frankreich wurden wesentlich mehr randomisierte Feldexperimente in der Sozialpolitikforschung durchgeführt als in Deutschland. Wichtige Bereiche und Instrumente der Sozialpolitik wurden in Deutschland bislang im Unterschied zu den USA und zu Frankreich nicht experimentell evaluiert. So ist der Zugang zu Sozialleistungen, beispielsweise zu Gesundheitsleistungen und zur sozialen Mindestsicherung, Gegenstand einer Reihe von Studien aus den USA. Auch in Deutschland ist die Frage, warum Sozialleistungen nicht in Anspruch genommen werden, eine Kernfrage der Sozialpolitik, doch gibt es keine einzige randomisierte Studie dazu.

Randomisierte Feldexperimente wären auch im Bereich der Grundsicherung, also des Bürgergelds, nützlich, zum Beispiel bei den Mitwirkungspflichten. Dies könnte an vorhandene Ansätze in Deutschland anknüpfen. So zeigt eine Studie von van den Berg et al. (2021), dass die experimentelle Untersuchung von Eingliederungsvereinbarungen wertvolle empirische Evidenz schaffen kann – gerade in einem Bereich, in dem häufig normativ und nicht empirisch gestützt argumentiert wird.

In Frankreich wurden Maßnahmen zur Unterstützung von Jugendlichen im Übergang von der Schule in den Beruf besonders häufig experimentell evaluiert. Ein solcher Fokus ist sinnvoll, da gerade hier quasi-experimentelle Ansätze oft nicht durchführbar sind und Ansätze, die auf der Annahme der bedingten Unabhängigkeit (siehe Abschnitt 1.2) beruhen,

aufgrund von Datenlücken nicht plausibel sind. Fördermaßnahmen in diesem Bereich werden auch deshalb bislang in Deutschland nur in relativ geringem Umfang kausal evaluiert. Infolge der größeren Anzahl und thematischen Breite der randomisierten Feldexperimente in den USA gibt es dort eine Wissensbasis mit vielfältigen Ergebnissen, die dann auch z. B. durch Meta-Studien zusammengefasst werden kann. In Deutschland fehlt es dagegen an Breite und Vielfalt. Ein weitaus stärkeres Gewicht hat in Deutschland dagegen die quasi-experimentelle Sozialpolitikforschung, wie die Bestandsaufnahme in Abschnitt 3 zeigt. Die Durchführung solcher Studien ist sinnvoll, denn viele sozialpolitische Fragestellungen eignen sich nicht für randomisierte Experimente. Dennoch scheint in der Aufteilung der Forschungsressourcen zwischen randomisierten und quasi-experimentellen Studien in Deutschland eine Unwucht zu bestehen.

In anderer Hinsicht hat die experimentelle Evaluation der Sozialpolitik in Deutschland dagegen einen Vorsprung vor anderen Ländern. Die randomisierten Studien in Frankreich berücksichtigen überwiegend „harte“ Indikatoren als Ergebnisgrößen – z. B. Noten oder Testergebnisse, Bildungsübergänge, den Erwerbsstatus oder das Einkommen. Tendenziell fehlt es an der Berücksichtigung einer Vielfalt unterschiedlicher Ergebnisdimensionen. Bei den meisten Interventionen sind jedoch unterschiedliche Outcomes relevant – kurzfristige und langfristige, individuelle und gesellschaftliche. Hier öffnet sich Spielraum für interdisziplinäre Kooperationen, weil innerhalb desselben Experiments sowohl für Psychologen, Sozialwissenschaftler als auch Ökonomen interessante Outcomes erhoben werden können.

Die randomisierten Wirkungsstudien bilden in allen Ländern einen fließenden Übergang zur Grundlagenforschung, in der beispielsweise psychologisch begründete Hypothesen geprüft und allgemeine Erkenntnisse über Verhalten (z. B. Bereitschaft zur Zahlung von Beiträgen, Diskriminierung) gewonnen werden. Diese Forschung sollte im Hinblick auf sozialpolitische Fragestellungen erweitert werden. Vor dem Hintergrund der Zielsetzung, Polarisierung und Konflikte zu überwinden, sollte auch die Frage der Akzeptanz von Sozialpolitik experimentell untersucht werden. Ausbaufähig ist in Deutschland die Finanzierung und Infrastruktur für die Durchführung randomisierter Feldexperimente. Die experimentelle Forschung braucht Finanzierung. Staatliche Forschungsförderung durch das BMBF oder die DFG wurde bislang nur für einzelne Projekte gewährt. Spezifische Förderlinien gab es dagegen nicht. Das Vorhandensein von Forschungsmitteln ist eine wesentliche Voraussetzung für die Vielzahl der experimentellen Studien in den USA. Auch am Beispiel Frankreich ist nachzuvollziehen, wie stark die Gewährung, aber auch die Kürzung von Forschungsmitteln die Forschungsaktivitäten beeinflusst.

Die experimentelle Forschung braucht ferner den Zugang und die Kooperation mit Akteuren, die die Experimente durchführen, also z. B. Bildungsträger. Daher ist bei der Umsetzung von randomisierten Feldexperimenten eine frühzeitige, systematische und strategische Zusammenarbeit zwischen Wissenschaft und Politik notwendig. Auch hier ist der Vergleich mit anderen Ländern sinnvoll. In Frankreich wurden randomisierte Feldexperimente insbesondere im Zeitraum von 2007 bis 2010 von der Politik initiiert und systematisch angewendet. Zugleich zeigt das Beispiel Frankreichs aber auch die Gefahr einer zu engen Anbindung an die Politik, wenn die politische Unterstützung wieder zurückgeht.

Nicht zuletzt fehlt es in Deutschland an praktischer und sozialwissenschaftlicher, ethischer, rechtlicher sowie wissenschaftstheoretischer Reflexion darüber, was randomisierte Feldexperimente im Bereich der Sozialpolitik leisten können. In dieser Studie konnte z. B. der rechtliche Rahmen nur oberflächlich behandelt werden, genauere rechtliche Studien wären hilfreich. Auch in diesem Bereich sind die Vergleichsländer weiter fortgeschritten (z. B. Gueron/Rolston 2013, Jatteau 2016, Oakley 1998).

## Literatur

Abrams, E., Libgober, J., List, J. A. (2020). Research registries: Facts, myths, and possible improvements (No. w27250). National Bureau of Economic Research.

Arntz, M., Thomsen, S.L. (2011). Crowding Out Informal Care? Evidence from a Field Experiment in Germany. *Oxford Bulletin of Economics and Statistics*, 73 (3), 398-427.

Banerjee, A.V., Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, Random House.

Bauer, T.K., Bender, S., Bonin, H. (2007). Dismissal Protection and Worker Flows in Small Establishments. *Economica*, 74 (296), 804-821.

Bauernschuster et al. 2018.

Bauernschuster, S., Driva, A., Hornung, E. (2020). Bismarck's Health Insurance and the Mortality Decline. *Journal of the European Economic Association*, 18 (5), 2561-2607.

Boockmann, B., Osiander, C., Stops, M. (2014). Vermittlungsstrategien und Arbeitsmarkterfolg: Evidenz aus kombinierten Prozess- und Befragungsdaten, *Journal for Labour Market Research*, 47(4), 341-360.

Boockmann, B., Zwick, T., Ammermüller, A., Maier, M. (2012). Do Hiring Subsidies Reduce Unemployment Among Older Workers? Evidence from Natural Experiments. *Journal of the European Economic Association*, 10 (4), 735-764.

Bruckmeier, K., Wigger, B. (2014). The Effects of Tuition Fees on Transition from High School to University in Germany. *Economics of Education Review* 41, 14-23.

Buck, S., McGee, J. (2015). *Why government needs more randomized controlled trials: Refuting the myths*. Laura and John Arnold Foundation.

Bundesregierung. 2021. Fortentwicklung der Evaluierungskonzeption der Bundesregierung. Beschluss des St-Ausschusses Bessere Rechtsetzung und Bürokratieabbau vom 26. November 2019.

Bundesregierung. O. J. Wirksam regieren. Mit Bürgern für Bürger. Abgerufen am 18.08.2023 von <https://www.bundesregierung.de/breg-de/themen/wirksam-regieren>.

Caliendo, M., Tübbicke, S. (2021). Design and Effectiveness of Start-Up Subsidies: Evidence from a Policy Reform in Germany. CEPA Discussion Papers 30, Universität Potsdam.

Card, D., Kluve, J., Weber, a. (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3), 894-931.

De Cler, A.-P. (2021). The Rise and Fall of a Gold Standard. The case of Randomized Controlled Trials within the Experimentation for Youth Fund. Master's Thesis, SciencePo, Paris.

Dustmann, C., Lindner, A., Schönberg, U., Umkehrer, M., vom Berge, P. (2020). Reallocation Effects of the Minimum Wage. *Quarterly Journal of Economics*, 137 (1), 267-328.

Fischer, M., Grewenig, E., Lergetporer, P., Werner, K. (2019). The E-word on the public acceptance of experiments. Discussion Paper, No. 219, Ludwig-Maximilians-Universität München und Humboldt-Universität zu Berlin, Collaborative Research Center Transregio 190 - Rationality and Competition, München und Berlin.

Fonds d'expérimentation pour la jeunesse (FEJ) (2019), Qui sommes-nous? <https://www.experimentation-fej.injep.fr/115-qui-sommes-nous.html>.

Geyer, J., Welteke, C. (2021). Closing Routes to Retirement for Women: How Do They Respond? *Journal of Human Resources*, 56 (1), 311-341.

Gueron, J. M., Rolston, H. (2013). *Fighting for reliable evidence*. Russell sage foundation.

Halla, M., Kahe, C., Sausgruber, R. (2021). Testing for Ethnic Discrimination in Outpatient Health Care: Evidence from a Field Experiment in Germany. IZA Discussion Paper 14886.

Haskins, R. (2017). Presidential address: Making federal social programs work. *Journal of Policy Analysis and Management*, 36(2), 276-302.

James, S., Bricet, R. (2021). Regard sur l'expérimentation sociale. Enquête auprès des porteurs de projets du Fonds d'expérimentation pour la jeunesse (FEJ), INJEP Notes et rapports/Note thématique.

Jamison, J. C. (2019). The entry of randomized assignment into the social sciences. *Journal of Causal Inference*, 7(1).

Jankowski, M., Tepe, M. (2022). Experimentelle Methoden. In: Wenzelburger, G., Zohlnhöfer, R. (Hrsg.): *Handbuch Policy-Forschung*. Springer VS. Zweite Edition, im Erscheinen.

Jatteau, A. (2016). *Faire preuve par le chiffre? Le cas des expérimentations aléatoires en économie*. Thèse de doctorat de l'Université Paris-Sarclay.

Kubbe, I. (2021). Experimente und experimentelle Forschungsdesigns. In: Wagemann, C. et al. (Hrsg.), *Handbuch Methoden der Politikwissenschaft*. Springer.

Mergele, L., Weber, M. (2020). Public employment services under decentralization: Evidence from a natural experiment. *Journal of Public Economics*, 182, 104113.

Oakley, A. (1998). Experimentation and social interventions: a forgotten but important history. *Bmj*, 317(7167), 1239-1242.

Office of Head Start. (o. J.) Data & Reports. Abgerufen am 18.08.2023 von <https://www.acf.hhs.gov/ohs/reports>.

Powers, E., Witmer, H.L. (1951). An experiment in the prevention of delinquency: the Cambridge-Somerville Youth Study. New York: Columbia University Press.

Saniter, N., Schnitzlein, D., Siedler, T, (2019). Occupational knowledge and educational mobility: Evidence from the introduction of job information centers. *Economics of Education Review*, 69 (C), 108-124.

Statistisches Bundesamt (2022). Arbeitshilfe zur Evaluierung von Regelungen der Bundesregierung. Statistisches Bundesamt, Wiesbaden.

Thorndike, E. L., Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, 8, 246-261.

van den Berg, G., Hofmann, B., Stephan, G., Uhlendorff, A. (2021). Mandatory Integration Agreements for Unemployed Job Seekers: A Randomized Controlled Field Experiment in Germany. IZA Discussion Paper 14026.

Wenner, Judith; Bozorgmehr, Kayvan; Duwendag, Stella; Rolke, Kristin; Razum, Oliver Wenner et al. 2020 Differences in realized access to healthcare among newly arrived refugees in Germany: results from a natural quasi-experiment. *BMC Public Health* (2020) 20:846

Ziefle, A., Gangl, M. (2014). Do Women Respond to Changes in Family Policy? A Quasi-Experimental Study of the Duration of Mothers' Employment Interruptions in Germany. *European Sociological Review*, 30 (5), 562–581.

### Über die Autor\*innen

**Yvette Bodry** ist Doktorandin am Lehrstuhl für Außenwirtschaft an der Universität Hohenheim.

**Prof. Dr. Bernhard Boockmann** ist wissenschaftlicher Direktor des Instituts für Angewandte Wirtschaftsforschung an der Universität Tübingen. Institut für Angewandte Wirtschaftsforschung e. V., Schaffhausenstraße 73, 72072 Tübingen

**Dr. Philipp Kugler** ist wissenschaftlicher Mitarbeiter am Institut für Angewandte Wirtschaftsforschung an der Universität Tübingen.

**Verena von Zitzewitz** war am Institut für Angewandte Wirtschaftsforschung an der Universität Tübingen tätig.

Die Autor\*innen bedanken sich bei eine\*r anonymen Gutachter\*in, Prof. Dr. Gesine Stephan sowie den Teilnehmenden eines Mini-Workshops am DIFIS am 2. März 2022 für viele nützliche Anregungen und Hinweise. Ferner bedanken sie sich für die finanzielle Unterstützung dieser Forschung durch das DIFIS.

Für die Inhalte der vorliegenden Publikation sind ausschließlich die Verfasser\*innen verantwortlich.

### Impressum

DIFIS - Deutsches Institut für Interdisziplinäre Sozialpolitikforschung

Direktorin: Prof. Dr. Ute Klammer (Universität Duisburg-Essen)

Stellv. Direktor: Prof. Dr. Frank Nullmeier (Universität Bremen)

Standort Duisburg: Institut Arbeit und Qualifikation (IAQ), Forsthausweg 2, 47057 Duisburg

Standort Bremen: SOCIUM Forschungszentrum Ungleichheit und Sozialpolitik,

Mary-Somerville-Straße 5, 28359 Bremen

Homepage: [www.difis.org](http://www.difis.org)

**Erscheinungsort und -datum:** Duisburg/Bremen, August 2023

**Inhaltliche Betreuung:** Dr. Rebecca Schrader

**Betreuung der Publikationsreihe:** Dr. Miruna Bacali

**Satz:** Simon Rickel

**ISSN:** 2748-7199