



Rigoreuse Wirkungsevaluierung

Methodische Voraussetzungen und kontextsensitive Gestaltung

Katharina Kaepfel
Senior Research & Policy Associate, J-PAL Europe

6. Juni 2024



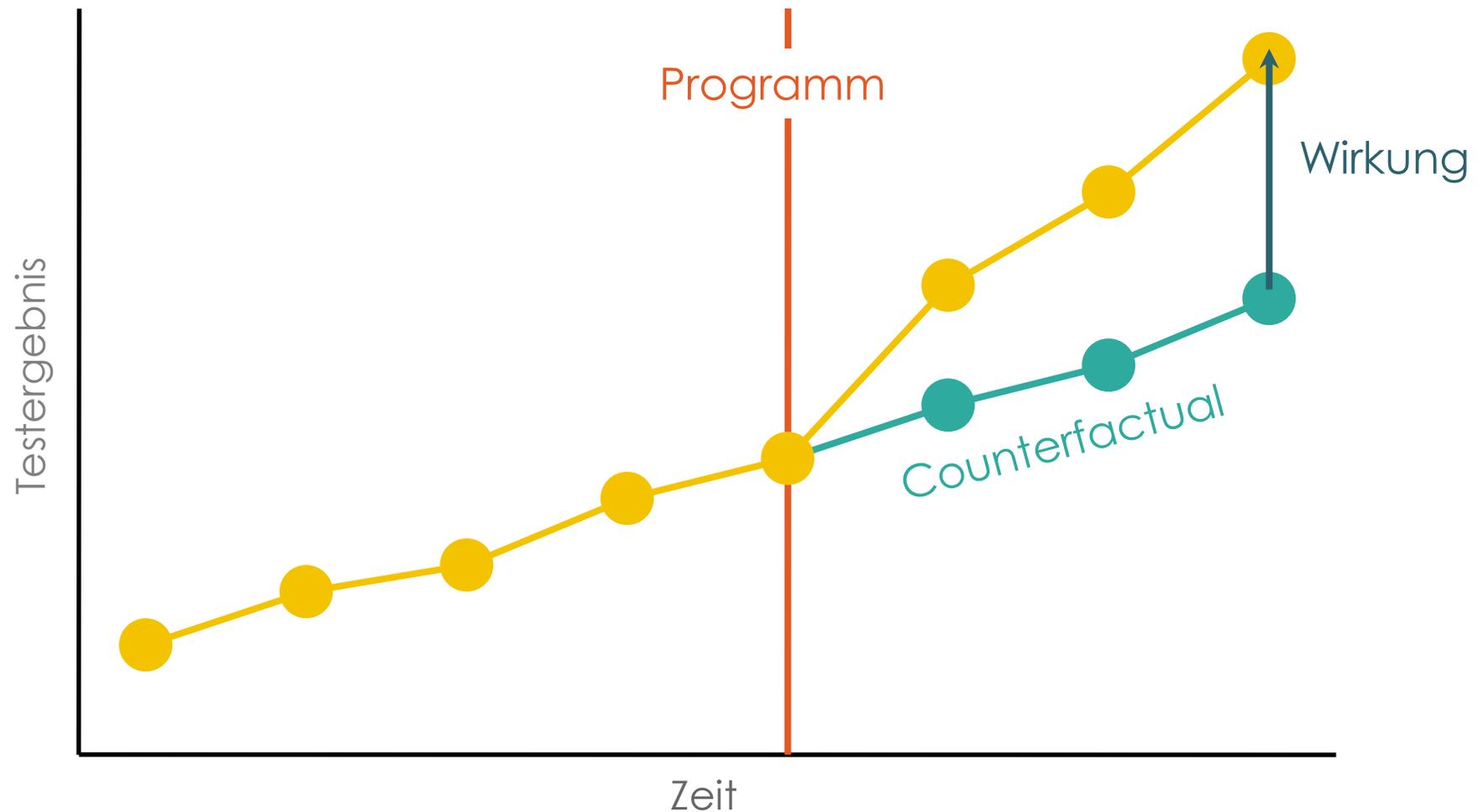
Inhalt

- I. Was verstehen wir unter Wirkung und wie kann man sie messen
- II. Beyond what works: Welche Fragen können randomisierte Evaluierungen beantworten?
- III. Beyond the method: Gute Programmplanung und -durchführung als Voraussetzung für Wirkungsmessung
- IV. Erwägungen, wann ein RCT sinnvoll sein kann

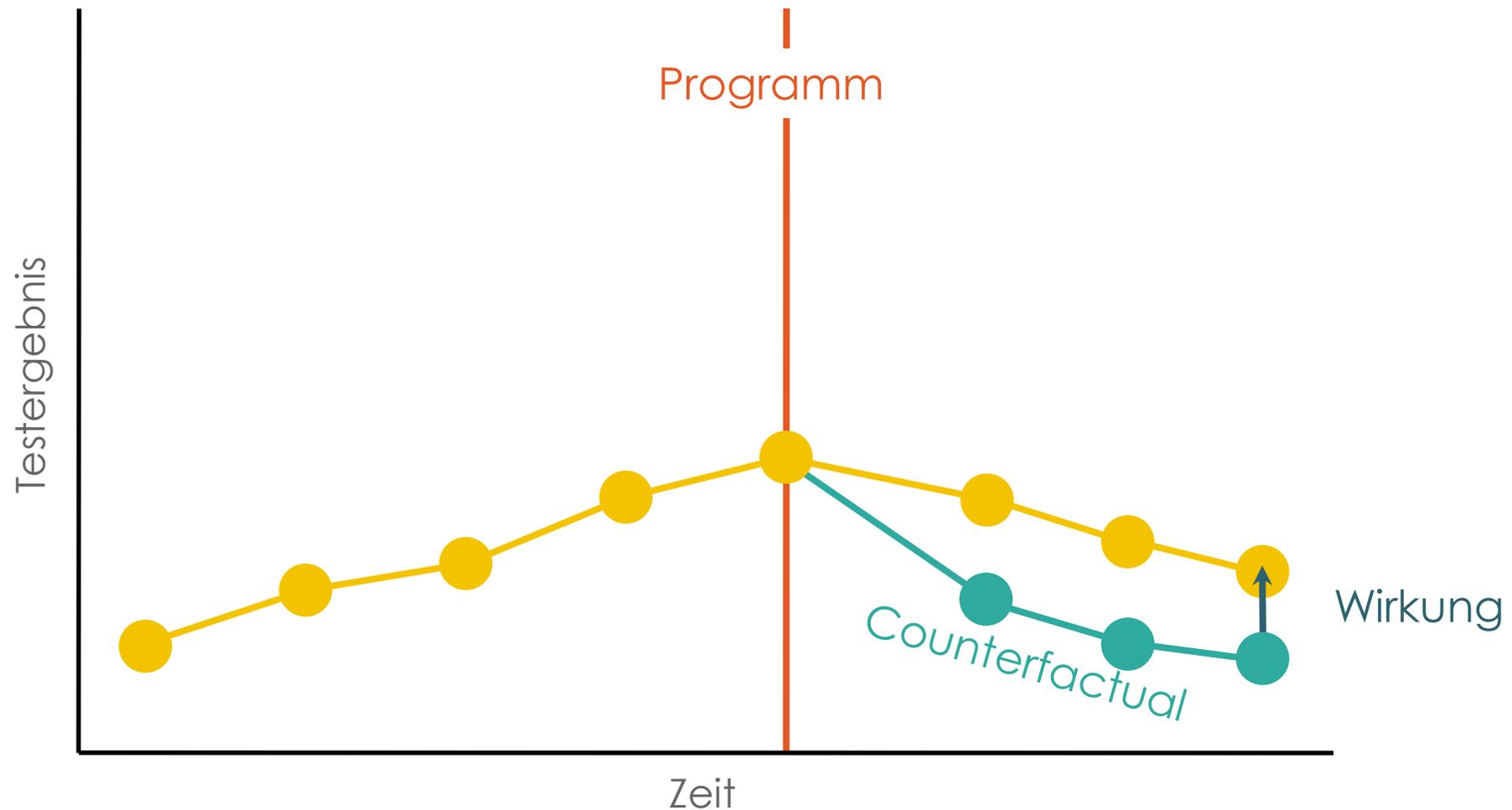
Inhalt

- I. **Was verstehen wir unter Wirkung und wie kann man sie messen**
- II. Beyond what works: Welche Fragen können randomisierte Evaluierungen beantworten?
- III. Beyond the method: Gute Programmplanung und -durchführung als Voraussetzung für Wirkungsmessung
- IV. Erwägungen, wann ein RCT sinnvoll sein kann

Was ist die Wirkung eines Nachhilfe-Programms auf Testergebnisse?



Was ist die Wirkung eines Nachhilfe-Programms auf Testergebnisse?



Um Wirkung zu messen, müssen wir das *Counterfactual* simulieren

Problem: Wir beobachten nicht, was in Abwesenheit des Programms passiert wäre.

Lösung: Counterfactual möglichst gut simulieren mit einer glaubwürdigen Kontrollgruppe

Versuchsgruppe



Source: freepik



Vergleichsgruppe

Nicht-experimentelle Methoden versuchen, *ex-post* eine gute Kontrollgruppe zu identifizieren

Nicht-experimentelle Methoden

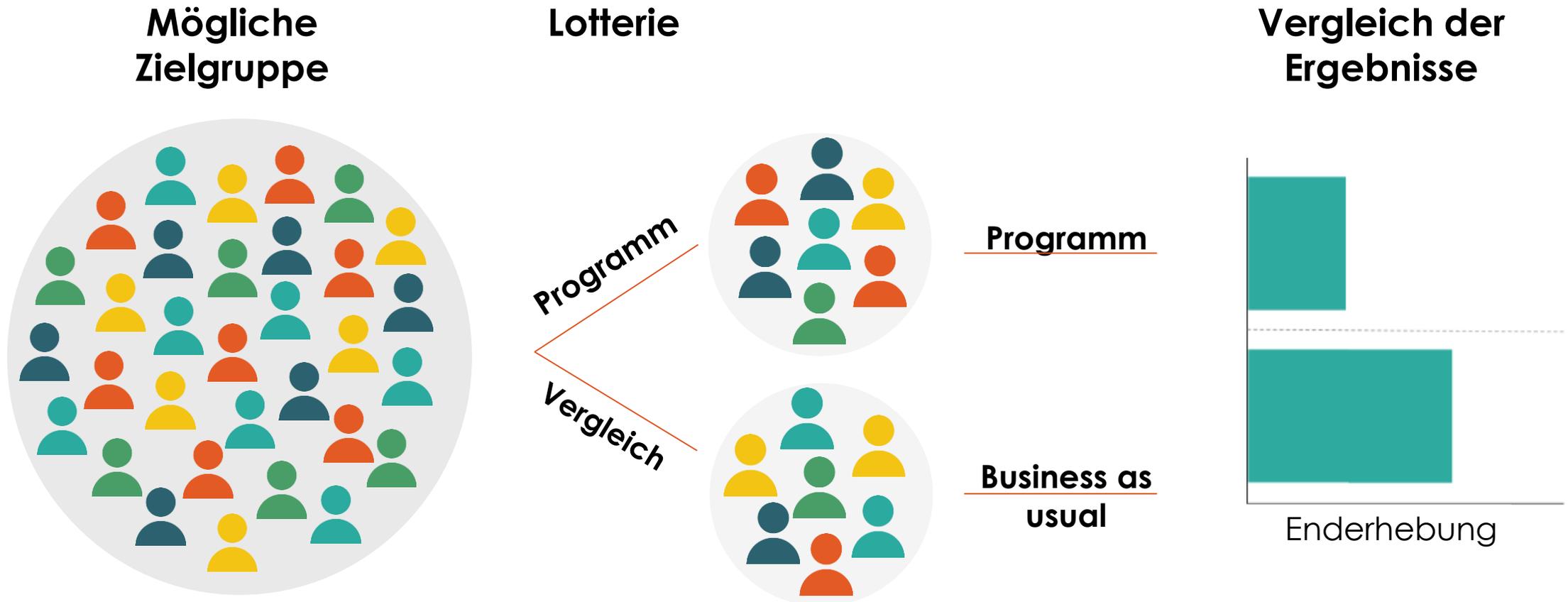
- Vorher-Nachher Vergleich
- Einfache Differenz
- Matching
- Differenz-in-Differenz
- Diskontinuitätsdesign
- Instrumentvariablen

Experimentelle Methode Randomisierte Evaluation

Auch bekannt als:

- Randomisierte Kontrollstudie (RCT)
- Experimente
- Soziale Experimente

RCTs nutzen eine Zufallsauswahl, um *ex-ante* vergleichbare Gruppen zu konstruieren



Inhalt

- I. Was verstehen wir unter Wirkung und wie kann man sie messen
- II. Beyond what works: Welche Fragen können randomisierte Evaluierungen beantworten?**
- III. Beyond the method: Gute Programmplanung und -durchführung als Voraussetzung für Wirkungsmessung
- IV. Erwägungen, wann ein RCT sinnvoll sein kann

RCTs können, je nach Erkenntnisinteresse, verschiedene Fragen über *kausale* Zusammenhänge beantworten

Wie effektiv ist ein Programm?

Wer profitiert am meisten?

Wie sind die verschiedenen Versionen eines Programms im Vergleich zueinander?

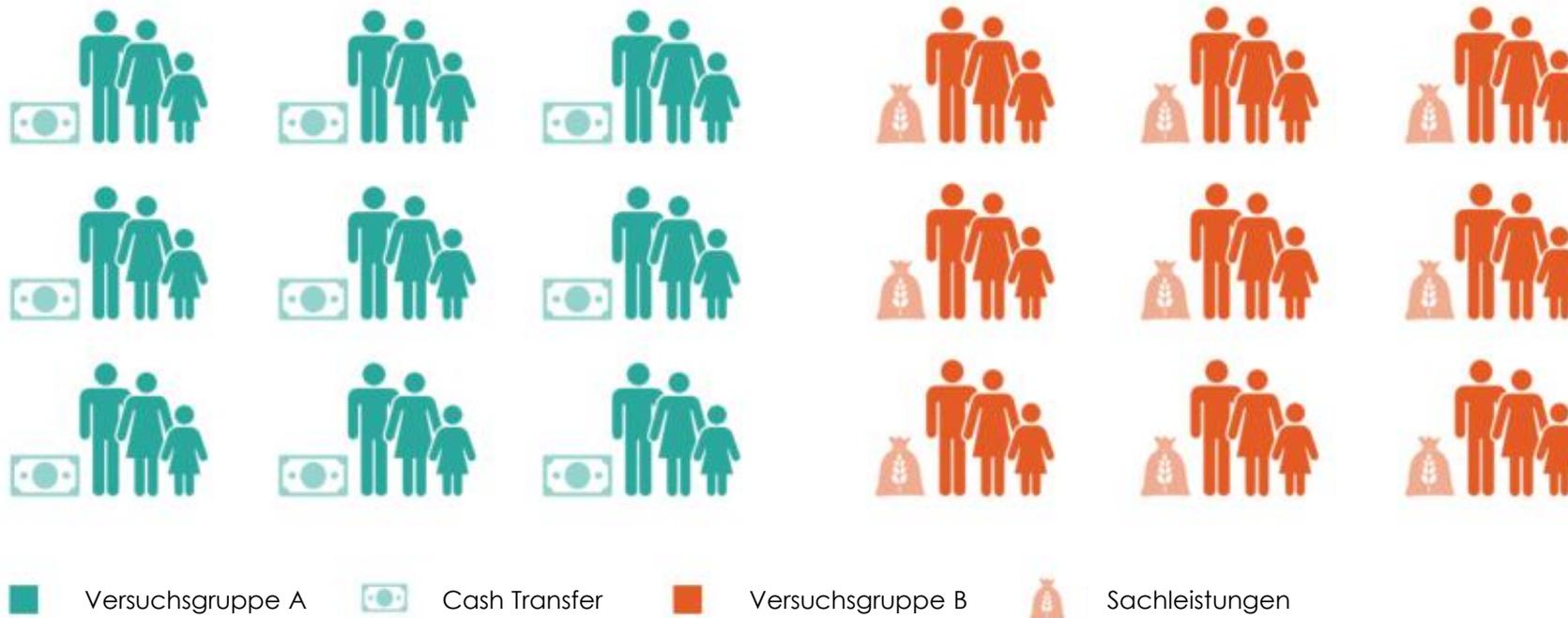
Gibt es Komplementaritäten?

Wie erreichen wir die, die Hilfe am meisten benötigen?

Wie erhöhen wir die Programmteilnahme?

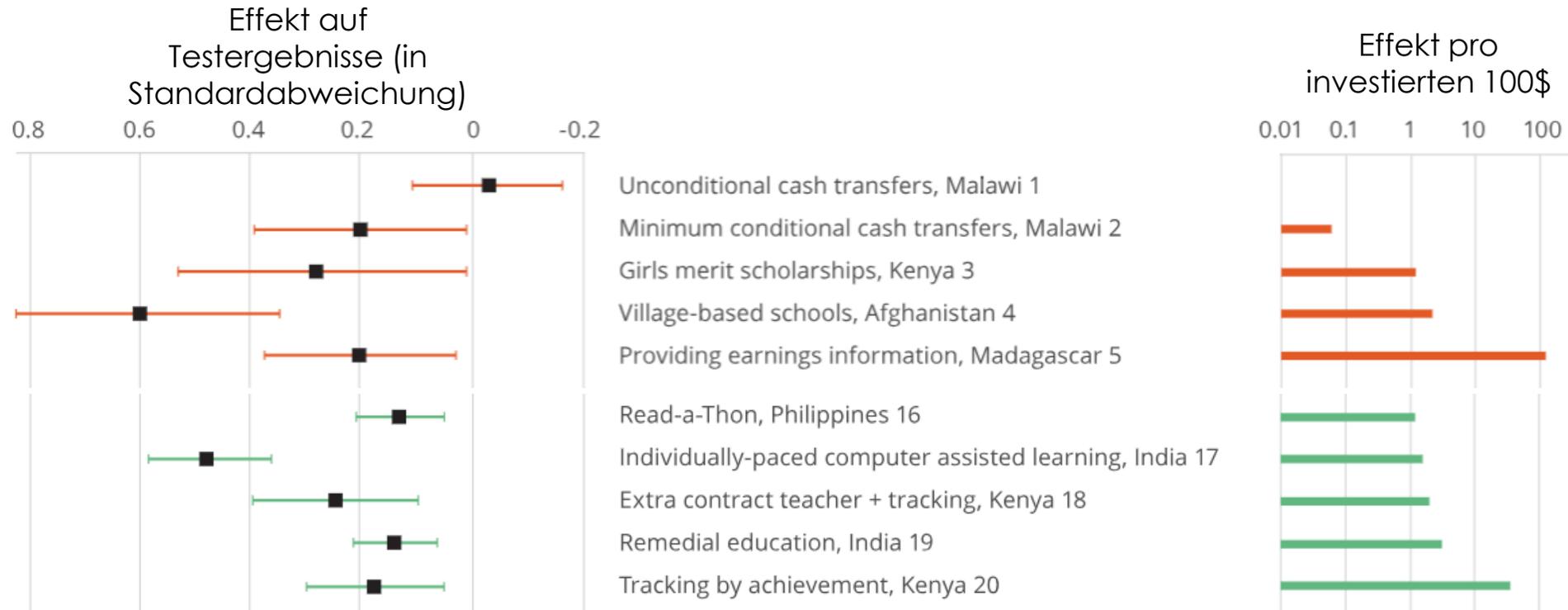
Welcher Ansatz bietet den größten Effekt für die gleichen Kosten?

Beispiel 1: Geld- vs. Sachleistungen



Aker ([2015](#)), Aker et al. ([2016](#))

Beispiel 2: Kosten-Wirkungs-Analyse

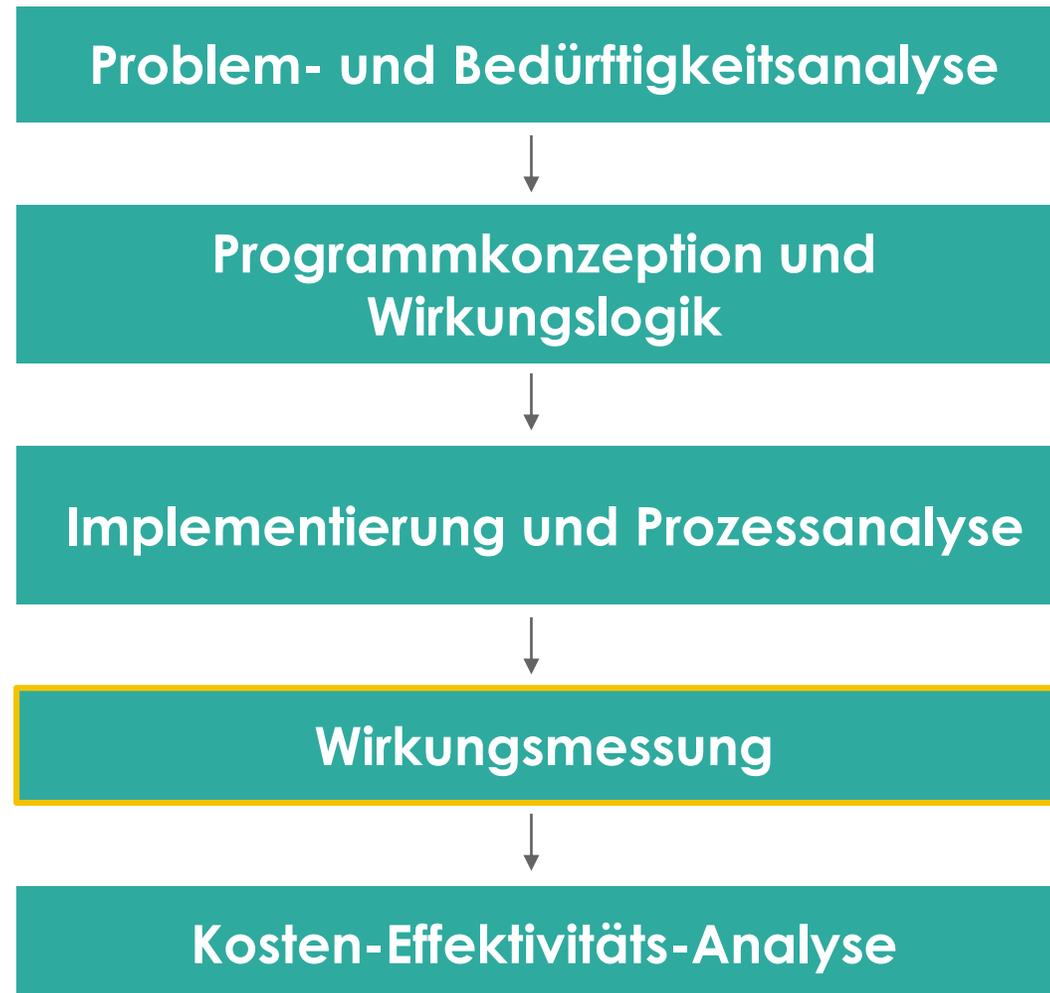


s. J-PAL's Ressource zur [Kostenwirksamkeitsanalyse](#)

Inhalt

- I. Was verstehen wir unter Wirkung und wie kann man sie messen
- II. Beyond what works: Welche Fragen können randomisierte Evaluierungen beantworten?
- III. Beyond the method: Gute Programmplanung und -durchführung als Voraussetzung für Wirkungsmessung**
- IV. Erwägungen, wann ein RCT sinnvoll sein kann

Eine gute Wirkungsevaluierung baut auf anderen Evaluationsansätzen auf



Inhalt

- I. Was verstehen wir unter Wirkung und wie kann man sie messen
- II. Beyond what works: Welche Fragen können randomisierte Evaluierungen beantworten?
- III. Beyond the method: Gute Programmplanung und -durchführung als Voraussetzung für Wirkungsmessung
- IV. Erwägungen, wann ein RCT sinnvoll sein kann**

Einige Fragen, bevor man eine Wirkungsevaluierung beginnt

Gibt es ein Element, das randomisiert werden kann?

Ist das Evaluationsdesign glaubhaft?

Ist das Programm “bereit”?
Braucht es vorerst Anpassungen?

Bestehen offene Fragen über die Wirksamkeit des Programms?

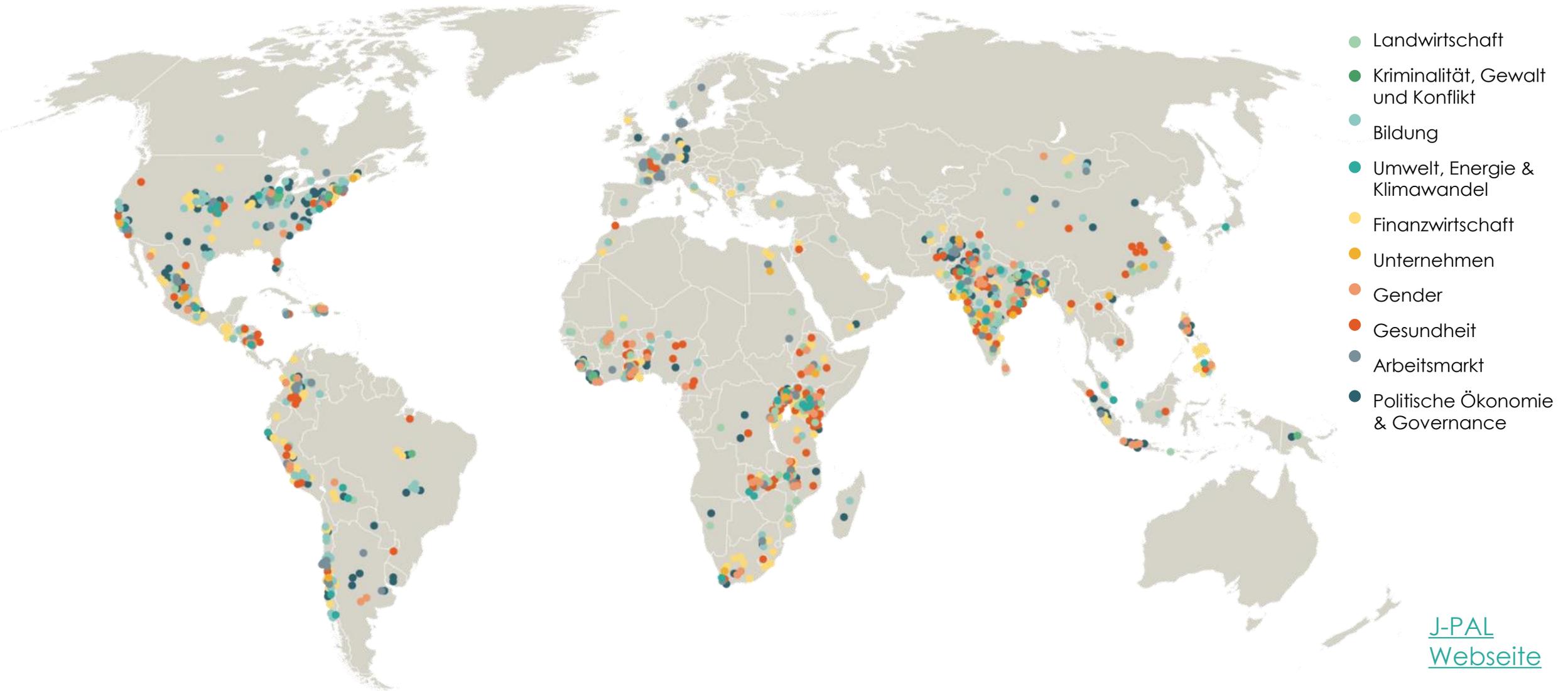
Besteht die Möglichkeit, Ergebnisse in konkrete Entscheidungen einfließen zu lassen?

Überwiegt der Nutzen der Evaluierung auf lange Sicht die Kosten?

Key Take Aways

- 1) Um die **Wirkung** eines Programms messen zu können, müssen wir das **Counterfactual** so gut wie möglich simulieren. Wenn sie richtig konzipiert und durchgeführt werden, können randomisierte Evaluierungen eine sehr glaubwürdige Methode sein, um die Wirkung eines Programms zu messen.
- 2) Wirkungsevaluierungen sind eines von vielen Tools. Wenn sie glaubhaft gestaltet sind, können sie verschiedene Fragen über **Ursache-Wirkungs-**Beziehungen beantworten.
- 3) Eine gute Wirkungsevaluierung baut auf einer guten Programmplanung und -durchführung auf. Verschiedene Evaluationsansätze **ergänzen** sich.

J-PAL's Forschung: 2,200+ Studien in 93+ Ländern



Vielen Dank!



Referenzen

Aker, J. C. (2017). [Comparing cash and voucher transfers in a humanitarian context: Evidence from the Democratic Republic of Congo](#). *The World Bank Economic Review*, 31(1), 44-70.

Aker, J. C., Boumnijel, R., McClelland, A., & Tierney, N. (2016). [Payment mechanisms and antipoverty programs: Evidence from a mobile money cash transfer experiment in Niger](#). *Economic Development and Cultural Change*, 65(1), 1-37.

Nützliche Ressourcen

- J-PAL One Pager: [Why randomise?](#)
- J-PAL Research Resource: [Introduction to randomized evaluations](#)
- J-PAL Research Resource: [The elements of a randomized evaluation](#)
- J-PAL Research Resource: [Assessing viability and building relationships](#)
- J-PAL's table of [Impact Evaluation Methods](#)
- J-PAL's [Advantages of Randomized Evaluations](#)
- J-PAL's [Common Questions and Concerns about Randomized Evaluations](#)
- J-PAL's [Generalisability Framework](#)
- J-PAL's Blog Post [Strengthening randomized evaluations through incorporating qualitative research, Part 1 \(Part 2, Part 3\)](#)