»Möglichkeiten und Grenzen der Wirkungsmessung«

Gesundheitsförderung Schweiz

18. Nationale Gesundheitsförderungs-Konferenz

19. Januar 2017, Neuenburg

Prof. Dr. Alexandra Caspari caspari@fb4.fra-uas.de
Fachbereich 4 Soziale Arbeit und Gesundheit

Was sind Wirkungen?

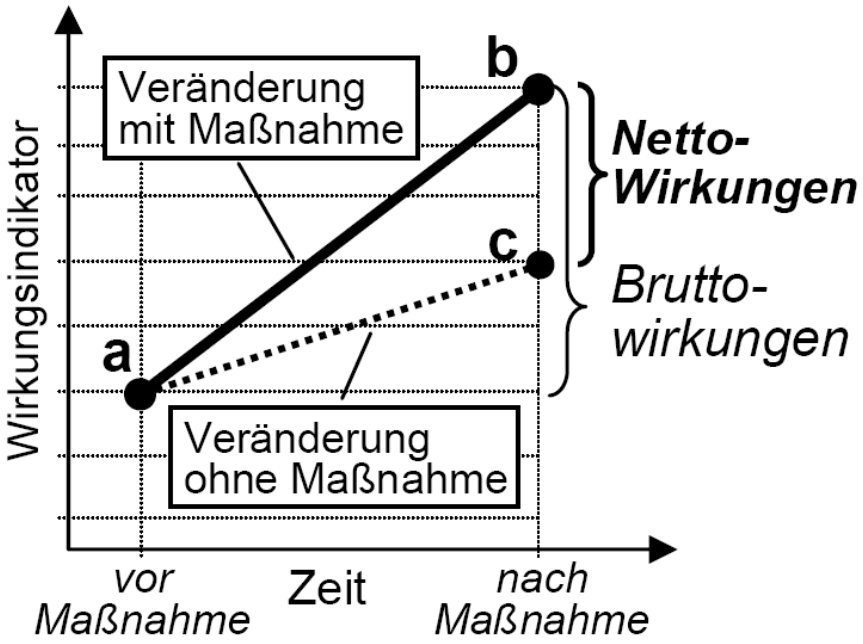
- Veränderungen nach Beendigung einer Maßnahme, die sowohl auf Maßnahme als auch beliebige Anzahl anderer Einflüsse zurückzuführen sind

  = ***Bruttowirkungen*** (Differenz b-a)

- Veränderungen nach Beendigung einer Maßnahme, die *allein* auf die durchgeführte Maßnahme zurückzuführen sind → isolierter Anteil an insgesamt auftretenden Veränderungen, die nicht beobachtbar gewesen wären, wenn Maßnahme nicht durchgeführt worden wäre

  = ***Nettowirkungen*** *oder* ***Projektwirkung*** (Differenz b-c)

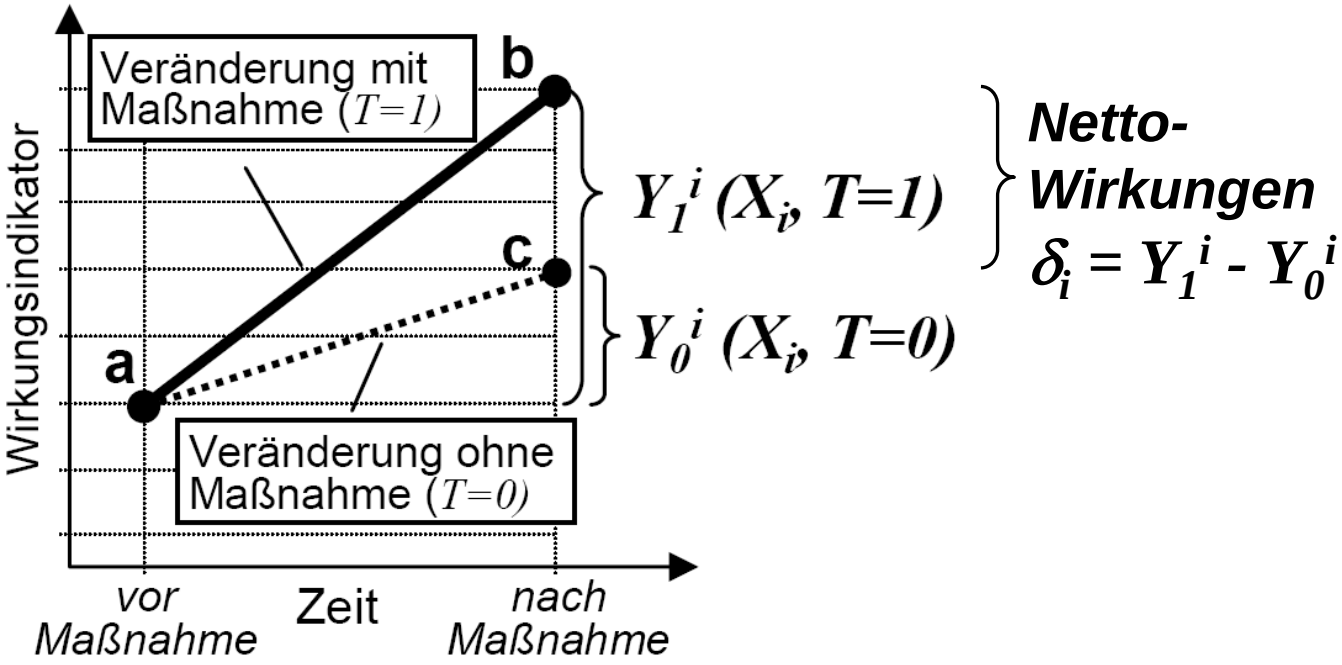
  = *kausaler Effekt*



Nettowirkungen/Kausaler Effekt

- Ein vom Auftreten eines kausal wirksamen Faktors T (Maßnahme) abhängiger kausaler Effekt δ_i (Wirkung) ist die Differenz zw. dem Ereignis Y_1^i, das bei Auftreten von $T\ (T=1)$ realisiert wird, und dem alternativen Ereignis Y_0^i, das ohne $T\ (T=0)$ eintreten würde:

$$\delta_i = Y_1^i\ (X_i, T=1) - Y_0^i\ (X_i, T=0) = Y_1^i - Y_0^i$$



Nettowirkungen/Kausaler Effekt

- Ein vom Auftreten eines kausal wirksamen Faktors T (Maßnahme) abhängiger kausaler Effekt δ_i (Wirkung) ist die Differenz zw. dem Ereignis Y_1^i, das bei Auftreten von $T\ (T=1)$ realisiert wird, und dem alternativen Ereignis Y_0^i, das ohne $T\ (T=0)$ eintreten würde:

$$\delta_i = Y_1^i\ (X_i, T=1) - Y_0^i\ (X_i, T=0) = Y_1^i - Y_0^i$$

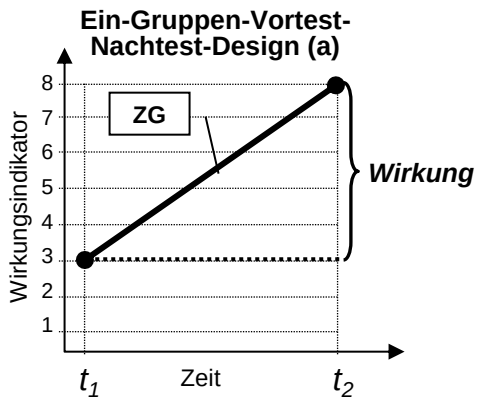
- Wirkungen sind nicht *direkt* beobachtbar:
  - Ereignis Y^i nur für $T=1\ (Y_1^i)$ *oder* $T=0\ (Y_0^i)$ beobachtbar
  - für Teilnehmer einer Maßnahme $(X_i, T=1)$ ist Ergebnis $Y_0^i\ (X_i, T=0)$ *nicht beobachtbar* (= ***das Kontrafaktische***)
- Wirkungen werden anhand *durchschnittlicher* Werte *empirisch erschlossen:* $\hat{\delta} = \overline{Y}_1 - \overline{Y}_0$
- Vergleich Ereignis bei Zielgruppe (ZG) und *hypothetischem Ereignis*, das ohne Maßnahme eingetreten wären

→ Frage des *Untersuchungsdesigns/Evaluationsdesigns*

vorexperimentelle Untersuchungsdesigns

- häufig genutztes Design zur (angeblichen) „Wirkungsmessung“:

	Design	Vorher-Messung t_1 (Baseline)	Stimulus	Nachher-Messung t_2 (Survey)
Vorexperimentelle/Nicht-experimentelle Versuchsachordnung:				
(a)	Ein-Gruppen-Vortest-Nachtest-Design	ZG_{t1}	X	ZG_{t2}
(b)	Ein-Gruppen-Nachtest-Design		X	ZG_{t2}



→ gemessene Wirkung = $\overline{Y}_{t2}^{ZG} - \overline{Y}_{t1}^{ZG}$

→ Kontrafaktische wird nicht berücksichtigt

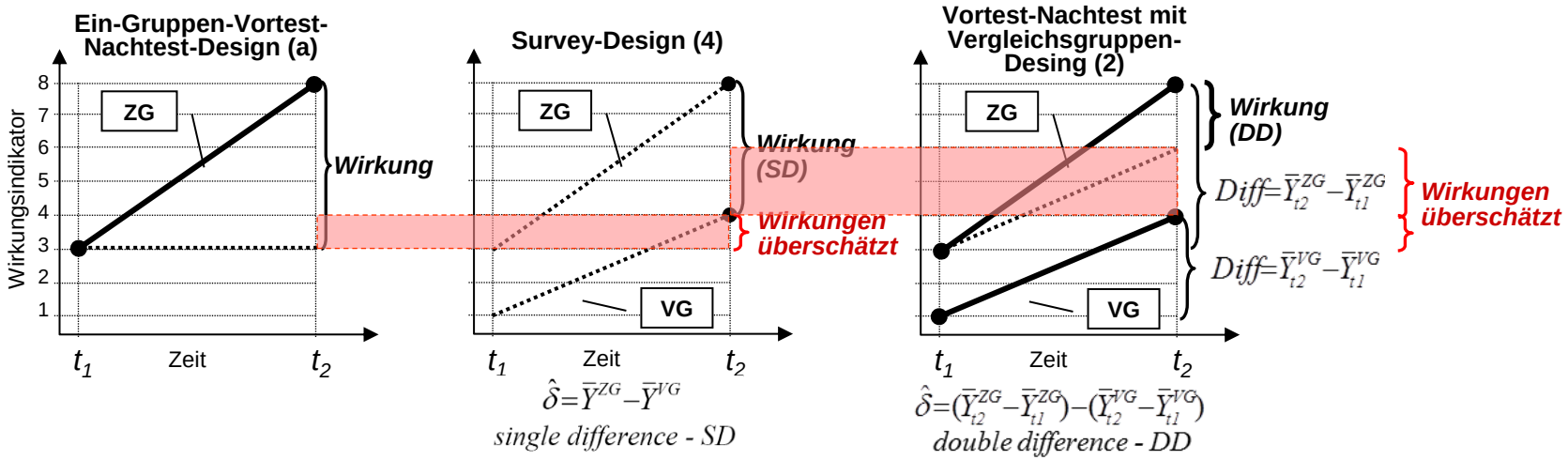
→ mit vorexperimentellen Designs kann Wirkung einer Maßnahme nicht nachgewiesen werden/keine kausale Attribution möglich

→ experimentelle bzw. quasi-experimentelle Designs notwendig

ZG: Zielgruppe, *X*: Stimulus (Projekt/Maßnahme), ● Datenerhebung, *t:* Zeitpunkt (erste, zweite Datenerhebung/Messung)

Berücksichtigung des Kontrafaktischen

Design	Vorher-Messung t_1 (Baseline)	Stimulus	Nachher-Messung t_2 (Survey)
Experimentelle Versuchsanordnung/"randomised controlled trial" (RCT):			
(1) Kontrollgruppen-Design	**ZG_{t1}**	**X**	**ZG_{t2}**
	KG_{t1}	**–**	**KG_{t2}**
Quasi-experimentelle Versuchsachordnung:			
(2) Vortest-Nachtest mit Vergleichsgruppen-Design	**ZG_{t1}**	**X**	**ZG_{t2}**
	VG_{t1}	**–**	**VG_{t2}**
(3) Vortest-Nachtest mit Nachtest Vergleichsgruppen-Design	**ZG_{t1}**	**X**	**ZG_{t2}**
		–	**VG_{t2}**
(4) Survey-Design		**X**	**ZG_{t2}**
		–	**VG_{t2}**



ZG: Zielgruppe, *VG:* Vergleichsgruppe, *X*: Stimulus (Projekt/Maßnahme), ● Datenerhebung, *t:* Zeitpunkt (erste, zweite Datenerhebung/Messung)

Experimentelle & Vorexperimentelle Designs

QUALITÄT

	Design	Vorher-Messung t_1 (Baseline)	Stimulus	Nachher-Messung t_2 (Survey)
Experimentelle Versuchsanordnung/"randomised controlled trial" (RCT):				
(1)	Kontrollgruppen-Design	ZG_{t1} KG_{t1}	X –	ZG_{t2} KG_{t2}
Quasi-experimentelle Versuchsachordnung:				
(2)	Vortest-Nachtest mit Vergleichsgruppen-Design	ZG_{t1} VG_{t1}	X –	ZG_{t2} VG_{t2}
(3)	Vortest-Nachtest mit Nachtest Vergleichsgruppen-Design	ZG_{t1}	X –	ZG_{t2} VG_{t2}
(4)	Survey-Design		X –	ZG_{t2} VG_{t2}
Vorexperimentelle/Nicht-experimentelle Versuchsachordnung:				
(a)	Ein-Gruppen-Vortest-Nachtest-Design	ZG_{t1}	X	ZG_{t2}
(b)	Ein-Gruppen-Nachtest-Design		X	ZG_{t2}

ZG: Zielgruppe, *VG:* Vergleichsgruppe, *X*: Stimulus (Projekt/Maßnahme), *t:* Zeitpunkt (erste, zweite Datenerhebung/Messung)

Umsetzungsmöglichkeiten

Erfahrungen aus der Evaluationspraxis:

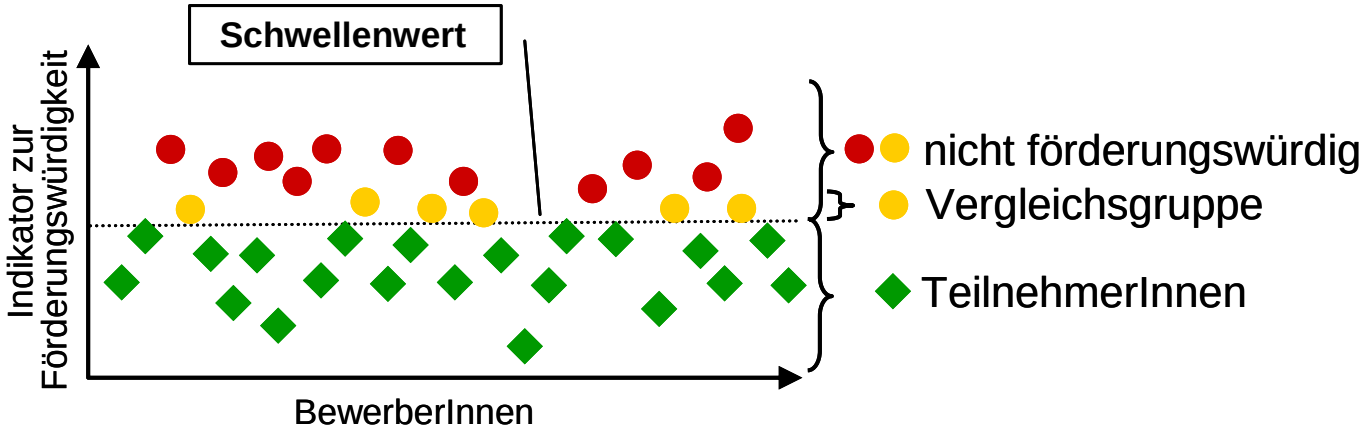
- angemessene Designs werden oft als unnötig anspruchsvoll, als methodisch unangemessen oder als nicht realisierbar abgelehnt
- Fokus hierbei meist „Goldstandard“ (RCTs)
- realistische Wege, wie *quasi-experimentelle* Designs in der Evaluationspraxis angewandt werden können, werden hierbei häufig nicht bedacht

Matching on Observables

- quasi-experimentelles Design:
  - bewusste Auswahl anhand gleicher charakteristischer Merkmale (relevanter Drittvariablen) der ZG
    z.B. Alter, Geschlecht, ökonomische Situation, etc.
  - VG wird aus Personen, Regionen, Bezirken gebildet, die höchste Übereinstimmung in Eigenschaften mit ZG aufweisen
  - nicht beobachtbare Merkmale („unobservables“), z.B. Motivation, schwer zu berücksichtigen
- Konstruktion einer VG für Nachher-Messung im Rahmen einer Evaluation möglich (t_2)
  → „nur“ single-difference (SD) möglich
- oder bereits bei Planung
  → auch Vorher-Messung (t_1) → double-difference (DD) möglich

Regression Discontinuity

- quasi-experimentelles Design: Konstruktion VG für Vorher- & Nachher-Messung
  - wenn Teilnahme an Maßnahme an bestimmte Voraussetzung mit gesetztem *Schwellenwert* gebunden, z.B. Einkommen, Alter, Testergebnis (BMI, Cholesterin), Leistung (Schulsport) etc.
  - wenn Erfüllung der Voraussetzung vorab überprüft wird
- VG = Personen, die Schwellenwert nur *knapp* nicht erreicht haben, aber sehr ähnliche Charakteristika wie ZG aufweisen

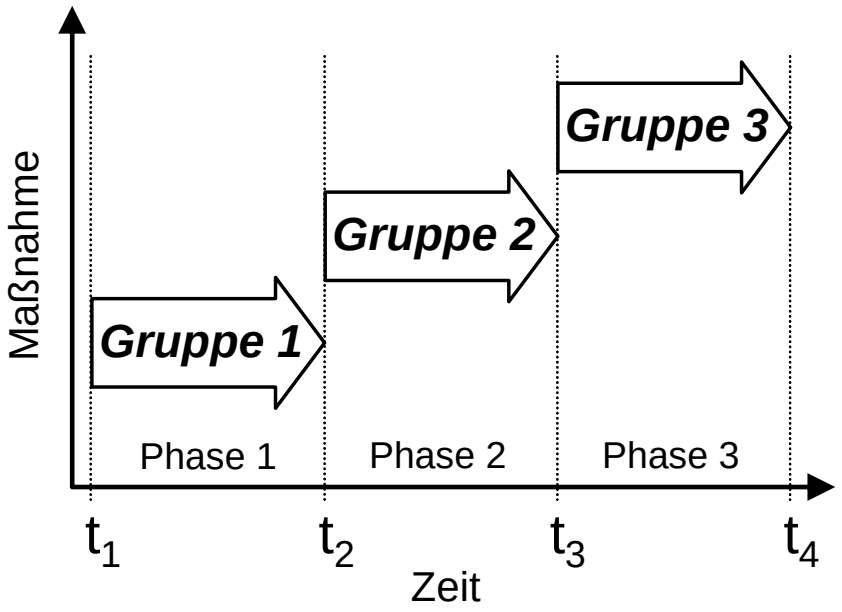


→ double-difference (DD) möglich

Pipeline Verfahren

- experimentelles Design: VG für Vorher- & Nachher-Messung
  - wenn größeres Programm mit langer Laufzeit in *mehreren Phasen zeitversetzt* implementiert wird (Schulen, Schulklassen, Städte, Stadtteile, Dörfer, Regionen)
  - wenn *keine bewusste Entscheidung* darüber, warum Klassen, Stadtteile, Dörfer etc. an der ersten Phase, andere erst später teilnehmen sollen („randomized phasing in“)
- Einheiten, die erst an der 2. & 3. Phase teilnehmen = VG für Personen der 1. Phase

→ double-difference (DD) möglich



Propensity Score Matching (PSM)

- quasi-experimentelles Design:
  - Konstruktion VG für Vorher- & Nachher-Messung
  - wenn Daten aus allgemeinen Surveys mit interessierenden Fragen zu Zeitpunkt t_1 und t_2 existieren
  - anhand charakteristischer Merkmale werden „Ähnlichkeitsindices“ geschätzt (berechnet)
  - auf Basis dieser „Ähnlichkeitsindices“ wird für jede Einheit der ZG eine (oder mehrere) „passende“ Einheiten aus dem Survey für VG ausgewählt, die sich bzgl. der Merkmale nicht von der ZG-Einheit unterscheidet („statistischer Zwilling“)

→ „Qualität“ der VG ~ KG

→ double-difference (DD) möglich

wichtige Anmerkung

- Internationale Diskussion um Wirkungsmessung bezieht sich nur auf *kleinen Ausschnitt* im Kontext einer Evaluation
  → Frage, wie *eindeutige Wirkungszuschreibung* (kausale Attribution) methodisch realisiert werden kann
  → Fokussiert auf *Outcomes* (direkte intendierte Wirkungen)
  = kleiner Ausschnitt des (komplexen) Wirkmodells/Wirkungsgefüges
- Daher: Wirkungsmessung ≠ Wirkungsevaluation!
- Wirkungsmessung *notwendig*, jedoch *nicht hinreichend*!
  → nur „Untersuchung“, ***ob*** Maßnahme wirkt *oder nicht*
  → Frage nach *Warum* bleibt unbeantwortet „Black Box“
- *aussagekräftige* Wirkungsevaluationen benötigen ebenso:
  → Analyse & Bewertung einer Maßnahme auch bzgl. anderer Fragen (Relevanz, Impact, Effizienz, Planung & Steuerung, etc.) auf allen Ebenen des Wirkmodells
  → Analyse nicht-intendierter Wirkungen

Möglichkeiten & Grenzen

- Wirkungsmessungen mit quasi-experimentellem Design sind verhältnismäßig einfach umzusetzen!
- Voraussetzung: Wirkmodell mit Indikatoren (größte Hürde)
- Möglichkeit eines quasi-experimentellen Designs sollte immer geprüft werden (ex-ante)
- Wichtig: Wirkungsmessung im Kontext von Evaluationen ist *eine* Methode von vielen (Methodenmix/Triangulation)!

Wirkungsmessung stärkt die Plausibilisierung der (Wirkungs-)Ergebnisse!

FRANKFURT UNIVERSITY OF APPLIED SCIENCES

Vielen Dank

für Ihre Aufmerksamkeit!