

# Willkommen zur Vorlesung Empirische Methoden I

## 4. Vorlesung: Forschungsdesigns

Prof. Dr. Wolfgang Ludwig-Mayerhofer

Universität Siegen – Philosophische Fakultät, Seminar für Sozialwissenschaften

# Was heißt Forschungsdesign?

- **Forschungsdesign im weiten Sinn:**  
Die Gesamtheit aller Entscheidungen über das Vorgehen bei einer empirischen Untersuchung.
- **Forschungsdesign im engeren Sinn** (vor allem in der standardisierten Sozialforschung):  
Die Entscheidung, ob man
  - ein ‚echtes‘ Experiment durchführen möchte oder nicht (und wenn ja, wie man dieses genau anlegt), sowie
  - bei nicht-experimentellen Studien ein Querschnitts-, Trend- oder Paneldesign wählt.

## Ausgangspunkt

Mit einiger Übertreibung, aber letztlich nicht ganz zu Unrecht kann man feststellen: **Das Standard-Design der empirischen Sozialforschung ist die Querschnittsstudie an Individuen** (typisch: Umfrageforschung; aber auch: Interviews in ‚qualitativer‘ Forschung).

Dies ist eine sehr starke Einschränkung. Zum Glück nimmt die Vielfalt von Untersuchungsformen zu. Zu den wichtigsten Innovationen der vergangenen 30 Jahre gehören Längsschnitt-, vor allem Paneluntersuchungen.

## Designs in der ‚qualitativen‘ Forschung

Der Begriff F. ist in der qualitativen Forschung wenig kanonisiert.

Er wird, wenn überhaupt, vor allem in der weiteren Bedeutung gebraucht. Selbstverständlich haben auch qualitative Forschungsprojekte ein Design – einen (revidierbaren) Plan, eine Sammlung von Vorstellungen, wie man welche Daten erheben und wie zu Schlussfolgerungen kommen will.

Zu vielen Fragen liegen noch zu wenig Erfahrungen vor (Längsschnittuntersuchungen mit Wiederholungsbefragungen, Experimente).

## Forschungsdesign (qualitativ) allgemein

Uwe Flick (Handbuch-Artikel, siehe Lit. am Ende) unterscheidet (sicher nicht vollständig)

- Einzelfall oder mehrere Fälle
  - Fallstudien
  - Vergleichsstudien
- Zeitdimension
  - retrospektive Studien
  - Momentaufnahmen
  - (,echte') Längsschnittstudien

## Forschungsdesign („qualitativ“) allgemein

Kriterien für / Einflüsse auf Entscheidungen sind (nach U. Flick)

- Zielsetzung der Untersuchung (Beschreibung, Hypothesenprüfung, Theoriebildung)
- Fragestellung
- Generalisierungsziele
- Darstellungsziele
- Grad an Standardisierung und Kontrolle
- Ressourcen

## Qualitative Forschung: Ein Beispiel

- Projekt: „Gemeinsam leben, getrennt wirtschaften? Grenzen der Individualisierung in Paarbeziehungen“
- Projektleitung: Jutta Allmendinger, Werner Schneider, Wolfgang Ludwig-Mayerhofer

## Beispiel: Definition und Auswahl der Fälle

- Was ist die Erhebungseinheit: Individuum oder Paar?
- Alle Paare oder nur Zweiverdienerpaare? (Hier: letzteres; Kontrastgruppe: Internationaler Vergleich)
- Wenige Paare, da explorativ; ausführliche Analysen
- Konkrete Fallauswahl (im ersten Schritt!)
  - Wegen Vergleichbarkeit tentativ bestimmte Kriterien (Dauer der Beziehung; mit/ohne Kind[er]; unterschiedliches Alter)
  - Einschränkungen: nur heterosexuelle Mittelschichtpaare mit homogenem kulturellen Hintergrund (also z. B. keine MigrantInnen)

## Beispiel: Quer- oder Längsschnitt?

- Paarbeziehungen ändern sich (vor allem in den ersten Jahren, nicht zuletzt wenn Kinder kommen), daher:
  - Längsschnittuntersuchung (Wiederholungsinterviews)
  - Möglichst Kohortendesign (dazu später mehr)
  - Wenn Paare sich trennen, Partner möglichst (auch in neuen Beziehungen) weiterverfolgen

## Beispiel: Welche Daten wie erheben?

- Wegen Bedeutung sinnhafter Paarkonstruktionen: Verbale Daten (Interviews); denkbar: Ergänzung durch Beobachtungsdaten (aber: darf man z. B. Menschen beim Einkaufen [genau, methodisch, wissenschaftlich] ohne Einverständnis beobachten?)
- Wie interviewt man Paare: Gemeinsam, getrennt oder beides? Wenn letzteres, in welcher Reihenfolge?
- Weitere Entscheidungen betreffen genaue Art der Datenerhebung, die Auswertung und die Präsentation der Daten

# Kausalität

Kausalität (Verursachung) ist schwer zu definieren. Für unsere Zwecke genügt es, Kausalität folgendermaßen aufzufassen:

- 1 Es muss ein Zusammenhang zwischen X und Y bestehen.
- 2 X (Ursache) muss Y (Wirkung) zeitlich vorausgehen. (Schließt nicht aus, dass Y auch auf X zurückwirkt).
- 3 Der Zusammenhang zwischen X und Y darf nicht durch andere Einflüsse (Z) bedingt sein.

Bedingung Nr. 3 ist der ‚Knackpunkt‘ !

# Wie sicher ist kausaler Einfluss?

- Je häufiger eine Person Gewaltdarstellungen in Medien betrachtet, desto häufiger neigt sie zu eigener Gewaltanwendung. [Einfluss der Medien?](#)
- Kranke Personen, die ein Medikament erhalten, fühlen sich nach zwei Wochen deutlich besser. [Einfluss des Medikaments?](#)
- Frauen erhalten (in Deutschland) nur knapp 70 Prozent des Lohns von Männern. [Folge von Lohndiskriminierung?](#)
- Arbeitslose sind häufiger krank als Nicht-Arbeitslose. [Folge der Arbeitslosigkeit?](#)

# Experiment (allgemein)

- Das Experiment gilt als der ‚Königsweg‘ zur Prüfung von Kausalbeziehungen. Es erlaubt die Feststellung von 1. (Zusammenhang zwischen X und Y) unter bestmöglicher Kontrolle von 2. (X vor Y) und 3. (Kontrolle von Z = anderen Einflüssen).
- Bekanntlich ist die Macht von Königen manchmal sehr gering. Bei allen Vorteilen von Experimenten bleibt eine große Schwäche: Ihre Verallgemeinerbarkeit auf nicht-experimentelle Gegebenheiten („externe Validität“) ist oft problematisch.

# Experiment (allgemein)

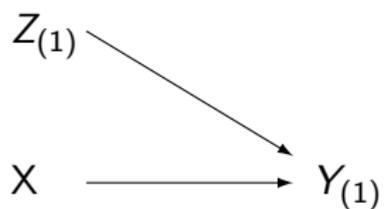
- Wir unterscheiden zwischen Feldexperimenten (Experiment in ‚natürlichen‘, vorhandenen Settings) und Laborexperimenten.
- Die Durchführung von Feldexperimenten z.B. zur Evaluation von Politikmaßnahmen ist in der BRD wenig verbreitet (im Gegensatz etwa zu den USA).
- Die Annahme, dass Experimente in Soziologie und Politikwissenschaft wenig sinnvoll sind, ist nur begrenzt richtig. Es ließen sich mit Sicherheit weitaus mehr Experimente durchführen als heute üblich.
- Ethische Probleme? Nicht, wenn Untersuchungsfrage wirklich offen ist.

# Kennzeichen von Experimenten

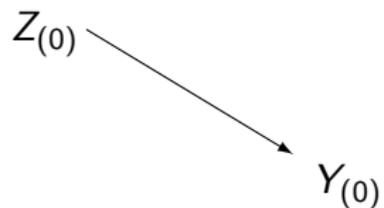
## Grundcharakteristika von Experimenten:

- Auf der Grundlage einer Hypothese wird ein Einfluss von X auf Y angenommen.
- Y ist (hinreichend gut) messbar.
- X („Stimulus“, „Treatment“) lässt sich kontrolliert setzen.
- „Störfaktoren“ (Z) werden kontrolliert durch
  - Elimination
  - Konstanthaltung oder
  - Randomisierung.

# Störfaktoren



Experimentalgruppe



Kontrollgruppe

# Störfaktoren: Elimination

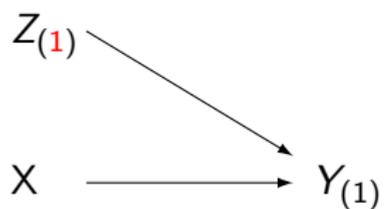
 ~~$Z_{(1)}$~~  $X \longrightarrow Y_{(1)}$ 

Experimentalgruppe

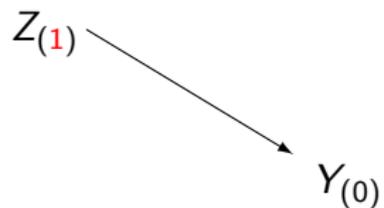
 ~~$Z_{(0)}$~~  $Y_{(0)}$ 

Kontrollgruppe

# Störfaktoren: Konstanthaltung

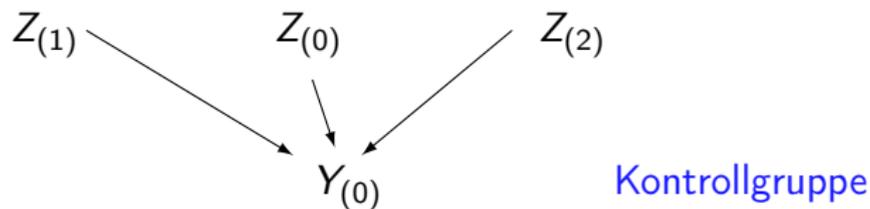
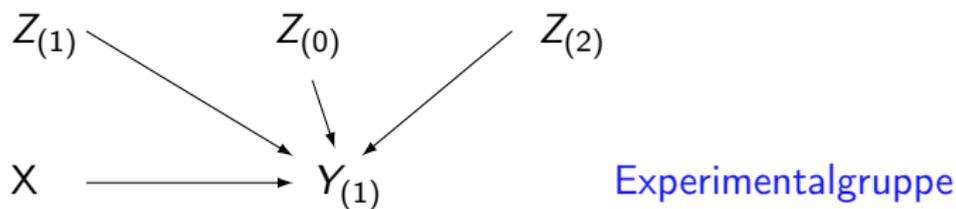


Experimentalgruppe



Kontrollgruppe

# Störfaktoren: Randomisierung



## Matching oder Randomisierung?

Verteilung der Untersuchungspersonen (-gruppen) auf E[xperimentalgruppe] oder K[ontrolgruppe] (evtl.  $E_1, E_2 \dots$ ):

**Matching:** zu jeder Person in E wird eine Person gesucht, die ihr in relevanten Merkmalen ähnelt, und K zugewiesen (moderne Variante: Propensity Score Matching).

Problem: Relevante Merkmale können unbekannt sein.

**Randomisierung:** Personen (Gruppen) werden **zufällig** E oder K zugewiesen.

Bei ausreichenden Fallzahlen sind im Schnitt Störgrößen gleich verteilt.

## Setzen des Treatments/Stimulus'

Vor allem in Feldexperimenten ist es schwierig, das gewünschte „Treatment“ zu isolieren:

- Wirkungen der Ganztagschule: Neue Unterrichtsformen, mehr (oder weniger) Enthusiasmus der LehrerInnen, Entlastung (oder Belastung) der Eltern, ...
- Wirkung von Tempo 30: Mehr Unachtsamkeit bei FußgängerInnen, mehr Ärger bei AutofahrerInnen, ...

# Praktische Probleme

- Welches Treatment soll geprüft werden?
  - Eine neue Unterrichtsmethode über eine Woche in einem Schulfach?
  - Die gleiche Methode über ein Jahr in allen (geeigneten) Fächern? (usw.)
- Wann (gegebenenfalls: wie oft) soll Erfolg gemessen werden, anders gesagt: Wann (und wie lange) soll Treatment wirken?
- Lässt sich X isolieren? Bsp.: Schüler und Lehrer aus verschiedenen Unterrichtsformen kommunizieren miteinander und beeinflussen sich wechselseitig.

## Experiment: Vorher-Nachher-Messung

$E$	$O_1$	$X$	$O_3$
$K$	$O_2$		$O_4$

$E/K$ : Experimental-/Kontrollgruppe;  $O$ : Messung („Observation“);  
 $X$  = Treatment, Stimulus

Hauptproblem: Effekt von  $O_1$  und  $O_2$  auf externe Validität (muss nicht zwingend auftreten, hängt von Art der Messung ab!)  
(technisch: Interaktion [Zusammenwirken] von  $O_1$  bzw.  $O_2$  und  $X/\text{non-}X$ ).

Beispiel: Ausfüllen eines Fragebogens zu Gewaltbereitschaft sensibilisiert Vpn für das Thema; Intelligenztest führt zu Lerneffekten hinsichtlich zweiten Tests.

## Experiment: Solomon-Vier-Gruppen-Design

$E_1$	$O_1$	$X$	$O_3$
$K_1$	$O_2$		$O_4$
$E_2$		$X$	$O_5$
$K_2$			$O_6$

Dieses Design erlaubt, die Wirkung von  $O_1$  bzw.  $O_2$  abzuschätzen.  
Es ist aber aufwändiger als Design 1.

# Experiment: Nur Nachher-Messung

$$E_1 \quad X \quad O_1$$
$$K_1 \quad O_2$$

Dieses Design ist im Grunde ausreichend, um (bei sonst richtiger Durchführung) aufzuzeigen, ob  $X$  eine Wirkung hat.

# Bedrohungen interner Validität I

**Interne Validität:** Gültigkeit der Annahme, dass die beobachteten Unterschiede zwischen den E und K tatsächlich auf X zurückzuführen sind.

- **History:** Zwischenzeitliches Geschehen „außerhalb“ des Experiments
- **Maturation:** Reifung = Entwicklung der Vpn
- **Testing:** Messeffekte = Effekte des Gemessen-Werdens (siehe oben: Problem für externe, aber nicht für interne Validität)
- **Instrumentation:** Instrumenteneffekte; Änderungen der Erhebungstechniken [oft notwendig!], der Untersuchungsleiter

## Bedrohungen interner Validität II

- **Selection**: Selbst- oder Fremdselektion
- (darunter:) **Statistical Regression**: Regression zum Mittelwert; besonders hohe und besonders niedrige Werte sind oft Ausreißer
- **Mortality, Attrition**: „Mortalität“ = Ausscheiden aus der Untersuchung
- **„Interne Reaktivität“** (mein Ausdruck!): der Effekt der Zuteilung zu E oder K (Kontrolle in medizinischen Studien: Blindstudie [Patienten wissen nicht, welche Behandlung sie bekommen] oder Doppelblindstudie [auch Behandler blind])

# Bedrohungen externer Validität

- Siehe oben: Problem von Design 1 (Auswirkung des Getestet-Werdens).
- **Interaktion von Selektion und X**: Wer nimmt an Experiment teil? (In Psychologie: Meistens Psychologie-Studierende!)
- **„Externe Reaktivität“**: Die Reaktion darauf, überhaupt (egal ob in E oder K) an einer Untersuchung/einem Experiment teilzunehmen.

## Hinweis: Quasi-Experimente

Unter Quasi-Experimenten (oder quasi-experimentellen Designs) versteht man Untersuchungsformen, die zwar nicht die gleiche interne Validität haben wie „echte“ Experimente, denen aber einige Aussagekraft zugestanden wird. Beispiele finden sich in der Literatur (Campbell & Stanley, Cook & Campbell, Shadish & Cook).

Im Rahmen dieser Einführungsvorlesung ist keine Zeit für eine Darstellung.

# Nicht-experimentelle Designs

Dabei sollte man versuchen (wenn erforderlich), den Kriterien für die Prüfung von Kausalität so gut wie möglich zu entsprechen:

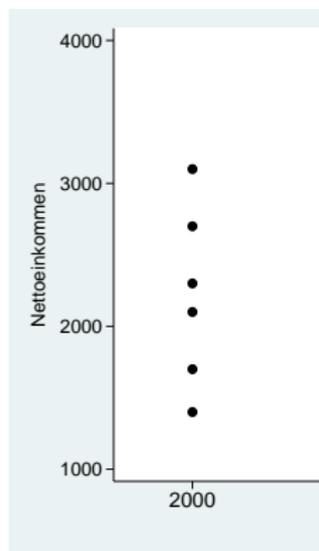
- ① (Zusammenhang zwischen X und Y) im Prinzip leicht feststellbar (aber: Messfehler!).
- ② (X vor Y) durch Wahl eines/r geeigneten Designs bzw. Erhebungsmethode auch im Prinzip in den Griff zu kriegen.
- ③ (Ausschaltung anderer Einflüsse) geschieht durch statistische Modellierung (tieferes Verständnis in Statistik III)

## Zeitliche Anordnung von X und Y

- Werden X und Y mit Bezug auf den gleichen Zeitpunkt erhoben, kann nicht mit Sicherheit entschieden werden, ob X Y beeinflusst oder umgekehrt (oder beides).
- In Längsschnittuntersuchungen (vor allem Untersuchungen mit Messwiederholungen [bei den gleichen Personen], Paneldesigns) lässt sich die zeitliche Ordnung von Merkmalen bestimmen, solche Untersuchungen können jedoch sehr aufwändig sein.
- Retrospektive Fragen sind teilweise auch über längere Zeiträume valide, teilweise (z. B. Einstellungen, Meinungen) jedoch unbrauchbar.

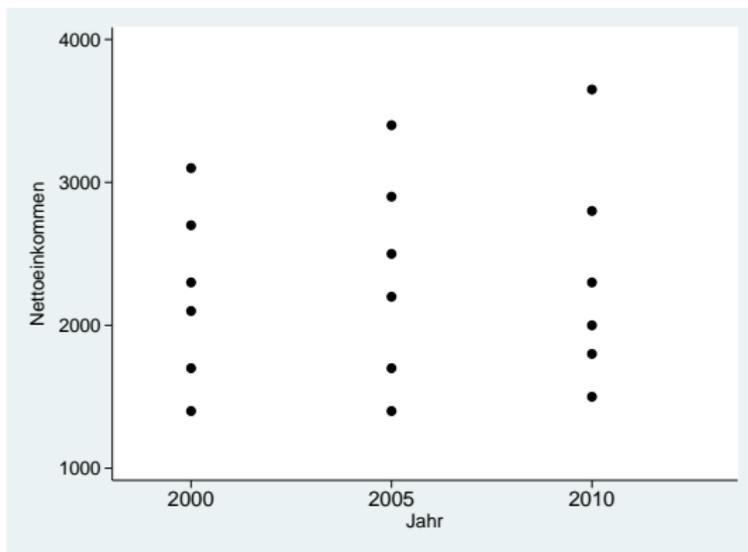
# Querschnittsstudie

Stichprobe aus allen Erwerbstätigen 2000 [fiktiv]



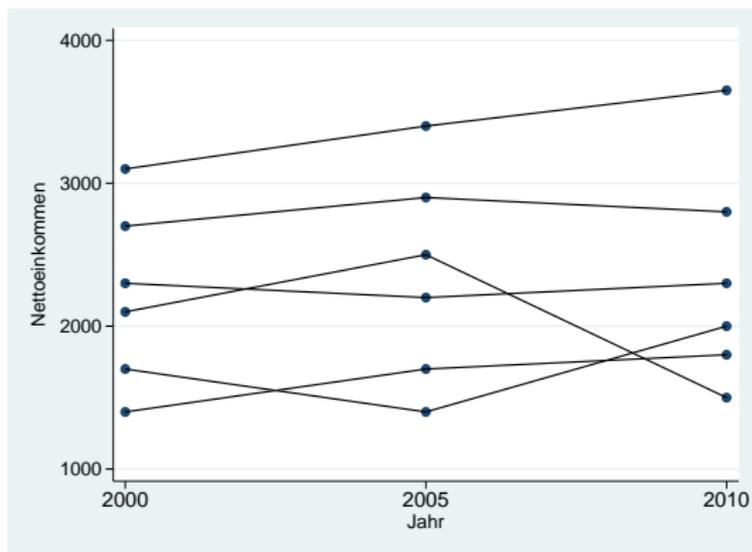
# Trendstudie

Stichproben aus allen Erwerbstätigen 2000, 2005, 2010 [fiktiv]



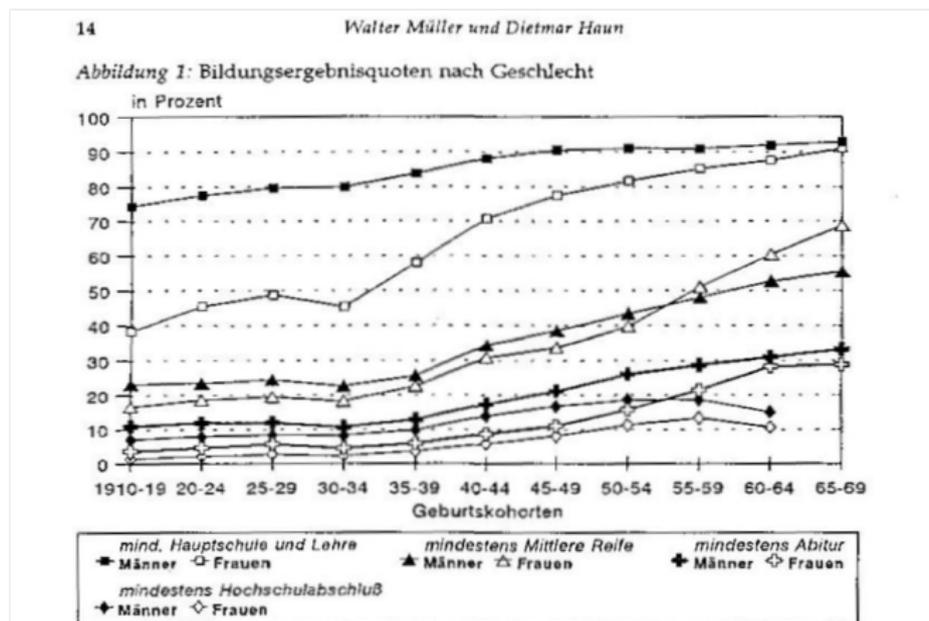
# Panelstudie

Stichprobe aus allen Erwerbstätigen 2000, Wiederholungsbefragung der gleichen Personen [fiktiv]



# Kohortenstudie

Gruppen (Kohorten) mit unterschiedlichen Geburtsjahren, Vergleich hinsichtlich gleichen Ereignisses [Quelle: Müller & Haun 1994]



## Zusätzliche Literatur

- Campbell, D. T., & J. C. Stanley, 1966, Experimental and Quasi-Experimental Designs for Research. Chicago: Rand McNally.
- Cook, T. D., & D. T. Campbell, 1979, Quasi-Experimentation. Design & Analysis Issues for Field Settings: Boston: Houghton Mifflin.
- Flick, U., 2000, Design und Prozess qualitativer Forschung, in U. Flick, E. v. Kardoff, und I. Steinke, Hrsg., Qualitative Forschung. Ein Handbuch: Reinbek bei Hamburg, Rowohlt, S. 252-265.
- Müller, W., & D. Haun, 1994, Bildungsungleichheit im sozialen Wandel: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 46, S. 1-42.
- Shadish, W. R., D. Cook & D. T. Campbell, 2001, Experimental and Quasi-experimental Designs for Generalised Causal Inference. Boston: Houghton Mifflin.