

# Observationsstudier I

Erik Gahner Larsen

Kausalanalyse i offentlig politik

# Seminaropgaven

- ▶ Send mig gerne hvad I har
  - ▶ Deadline 31. marts
- ▶ Tænk ikke så meget over antal sider
- ▶ Vejledning
- ▶ Dato for aflevering: ukendt

# Gæsteforelæsning

- ▶ Evalueringen af offentlige politikker
- ▶ Mandag den 18. april

# Dagsorden

- ▶ Forskellige tilgange
  - ▶ Eksperimenter, naturlige eksperimenter, observationsstudier
- ▶ Observationsstudier
- ▶ Forskellige designs
- ▶ Få cases
- ▶ Afbrudt tidsseriedesign

# Forrige gang

- ▶ Naturlige eksperimenter
  - ▶ RDD
  - ▶ IV
  - ▶ DID
- ▶ I dag bevæger vi os endnu længere væk fra eksperimenter

# I dag: Observationsstudier

- ▶ Observationsstudier er studier hvor forskeren
  - ▶ ikke kontrollerer datagenerationsprocessen og
  - ▶ der ikke er randomiseret variation
- ▶ Vores definition af kausalitet er dog den samme
  - ▶ “Like it or not, social scientists rely on the logic of experimentation even when analyzing nonexperimental data.” (Green og Gerber 2003, 110)
- ▶ Logikken er den samme
  - ▶ “In some sense every empirical researcher is reporting the results of an experiment. Every researcher who behaves as if an exogenous variable varies independently of an error term effectively views their data as coming from an experiment.” (Harrison og List 2004, 1009)

# Randomiserede og ikke-randomiserede grupper

- ▶ Randomiserede grupper og stimuli

R	X	$O_1$
R		$O_2$

- ▶ Ikke-randomiserede grupper og stimuli

NR	X	$O_1$
NR		$O_2$

- ▶ Randomiserede grupper faciliterer en kausal test
  - ▶ Ikke-randomiserede grupper kan **aldrig** facilitere en direkte kausal test
- ▶ Manglende randomisering: som regel overestimering af effekter (Kunz og Oxman 1998)

# Kvalitativ vs. kvantitativ metode

- ▶ Ikke en unik kvantitativ eller kvalitativ metode
- ▶ Alle kvalitative data kan måles og kodes med brugen af kvantitative tilgange
  - ▶ Og der er et komponent af kvalitativ forskning til al forskning
- ▶ Casestudier beror på samme logik som andre designs
  - ▶ “[T]here is no reason to suppose that case study research follows a divergent logic of inquiry relative to experimental research.” (Gerring og McDermott 2007, 689)
- ▶ Casestudier er en slags kvasiekperiment
  - ▶ Cases vælges med henblik på at teste et teoretisk argument



# Forskellige typer af studier (Blom-Hansen og Serritzlew 2014)

*Tabel 1:* Forskellige typer eksperimenter

	Sammenligning af eksperiment- og kontrolgrupper?	Interventionen er eksogen eller as-if eksogen	Inddeling i eksperiment- og kontrolgrupper er randomiseret eller as-if randomiseret?	Forskeren manipulerer interventionen?
Laboratorieeksperiment	Ja	Ja	Ja	Ja
Feltekperiment	Ja	Ja	Ja	Ja
Surveyeksperiment	Ja	Ja	Ja	Ja
Naturligt eksperiment	Ja	Ja	Ja	Nej
Kvasieksperiment	Ja	Ja	Nej	Nej
Traditionelt stort-N observationsstudie	Ja	Nej	Nej	Nej
Traditionelt singlecasestudie	Nej	Nej	Nej	Nej

# Designs

- ▶ Forskellige elementer kan styrke et observationsstudie
- ▶ Stimulitildeling
  - ▶ Tilfældig
  - ▶ Tærskelværdibaseret
  - ▶ Matching

# Designs

- ▶ Mål
  - ▶ Posttest observationer
  - ▶ Pretest observationer
  - ▶ Moderatorvariable
  - ▶ Måle på validitetstrusler

# Designs

- ▶ Sammenlignelige grupper
  - ▶ To eller flere forskellige grupper
  - ▶ Kohorter
  - ▶ Intern vs. ekstern kontrol
  - ▶ Konstruerede kontraster

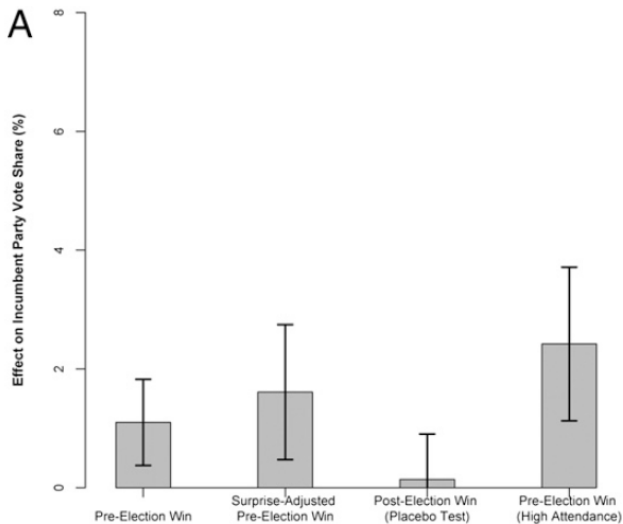
# Designs

- ▶ Stimuli
  - ▶ Skiftende stimuli
  - ▶ Modsatte stimuli
  - ▶ Fjernede stimuli
  - ▶ Gentagende stimuli

# Designfokus

- ▶ Fokus på designvalg (Rosenbaum 1999)
  - ▶ Valg af forskningsspørgsmål
  - ▶ Valg af grupper
  - ▶ Valg af konkurrerende teorier
  - ▶ Intern replikation
  - ▶ Fravær af stimuli i kontrolgruppen
  - ▶ Styrken af stimuli
  - ▶ *Natural blocks* (relaterede observationer)

# Eksempel: Effekten af college football på partivalg (Healy et al. 2010)



(a) Study 1: Effect of College Football Outcomes on Incumbent Party Vote Share, 1964-2006

# Eksempel: Effekten af college football på partivalg

- ▶ College football påvirker evalueringer af politikerne
  - ▶ Hvorfor?
- ▶ Observationsstudie
  - ▶ Vi kan ikke replikere det
- ▶ Hvordan kan vi styrke studiet?



## Eksempel: Effekten af college football på partivalg (Fowler og Montagnes 2015)

- ▶ Test teoretiske implikationer
- ▶ “In this study, we reassess the evidence on college football games and elections. Multiple independent sources of evidence suggest that this particular result is a false positive, despite the fact that Healy et al. (1) used a credible research design and followed the best practices of social science. There is no single smoking gun. Rather, we test several additional hypotheses that should hold if college football games truly influence elections, and in each case, the evidence fails to support the original finding.”

# Eksempel: Effekten af college football på partivalg (Fowler og Montagnes 2015)

Table 1. College football and incumbent support

	Baseline	Within county-year	By interest	By incumbent running	By home county	NFL
IncParty	.058* (.005)	.056* (.007)	.044* (.004)	-.006 (.007)	.057* (.001)	.043* (.004)
Football	-.008 (.008)		.000 (.010)	-.013 (.012)	-.005* (.001)	-.003 (.008)
IncParty × Football	.015 <sup>†</sup> (.007)	.003 (.011)	.026* (.008)	.021 (.012)	.011* (.001)	.004 (0.007)
IncParty × High Interest			.036* (.010)			
Football × High Interest			-.019 (.016)			
IncParty × Football × High Interest			-.029 <sup>†</sup> (.014)			
IncRunning				-.019 <sup>†</sup> (.009)		
IncParty × IncRunning				.092* (.008)		
Football × IncRunning				.007 (.013)		
IncParty × Football × IncRunning				-.007 (.013)		
Home County					.002 (.005)	
IncParty × Home County					-.007 (.005)	
Football × Home County					-.003 (.009)	
IncParty × Football × Home County					.002 (.007)	
Year-office fixed effects	X	X	X	X	X	X
County fixed effects	X		X	X	X	X
County-year fixed effects		X				
Observations	2,408	2,408	2,408	2,408	227,936	905
R <sup>2</sup>	0.472	0.832	0.478	0.516	0.610	0.640

County-clustered SEs are in parentheses; details are in the text. Substantively relevant coefficients are formatted in bold.

\*P < 0.01.

<sup>†</sup>P < 0.05.

# Designs uden kontrolgrupper

- ▶ En gruppe med posttest
  - ▶ Ingen pretests eller kontrolgruppe
  - ▶ Alle trusler i forhold til intern validitet
    - ▶ Med undtagelse af tidsdimensionen

$$\overline{X \quad O_1}$$

# Designs uden kontrolgrupper

- ▶ En gruppe med forskellige posttests
  - ▶ Ingen pretests eller kontrolgruppe
  - ▶ Samme trusler til intern validitet som med én posttest

$$\overline{X \quad \{O_{1A}O_{1B}\dots O_{1N}\}}$$

# Designs uden kontrolgrupper

- ▶ En gruppe med pretest og posttest

$$\overline{O_1 \quad X \quad O_2}$$

- ▶ En styrke at have en pretest
  - ▶ Men mangler stadig kontrolgruppe!
  - ▶ Intet kontrafaktisk scenarie

# Designs uden kontrolgrupper

- ▶ En gruppe med flere pretests og en posttest

$$\frac{O_1 \quad O_2 \quad X \quad O_3}{\quad}$$

- ▶ Giver mulighed for at undersøge modningseffekter
  - ▶ Men tager ikke nødvendigvis højde for historieeffekter, instrumenteffekter (og interaktioner mellem dem - eller interaktioner med modningseffekter)

# Designs uden kontrolgrupper

- ▶ En gruppe med pretest og posttest ved brug af forskellige afhængige variable

$$\overline{\{O_{1A}, O_{1B}\} \quad X \quad \{O_{2A}, O_{2B}\}}$$

- ▶ A forventes at blive påvirket af stimuli, B gør ikke
- ▶ B forventes at reagere på truslerne til intern validitet på samme måde som A

# Designs uden kontrolgrupper

- ▶ En gruppe med og uden stimuli

$O_1$	$X$	$O_2$	$O_3$	$X'$	$O_4$
-------	-----	-------	-------	------	-------

- ▶  $X'$  angiver at fjerne et stimuli
  - ▶ Tænk på en stikkontakt
- ▶ Kan bruges til at teste forskelle som funktion af stimuli i forhold til forskellige validitetstrusler
  - ▶  $X'$  skal ikke have samme effekt som  $X$ !



# Designs uden kontrolgrupper

- ▶ En gruppe med gentaget stimuli

$O_1$   $X$   $O_2$   $X'$   $O_3$   $X$   $O_4$

- ▶ Når der er mulighed for at introducere stimuli igen
  - ▶ Tænk (igen) på en stikkontakt

# Designs med kontrolgrupper uden randomisering

- ▶ Posttest med forskellige grupper

NR	$X$	$O_1$
NR		$O_2$

- ▶ Udfordring: Vi kan ikke sige om det er  $X$  eller forskelle mellem grupperne, der forklarer en eventuel forskel mellem  $O_1$  og  $O_2$
- ▶ Løsning: Forsøge at gøre de to grupper sammenlignelige, så man kan estimere effekten af  $X$ 
  - ▶ Næste gang ser vi på en af disse teknikker

# Designs med ikke-randomiserede kontrolgrupper og pre-tests

- ▶ To grupper med pre- og post-tests

NR	$O_1$	$X$	$O_2$
NR	$O_1$		$O_2$

- ▶ Difference-in-difference (jvf. sidste uge)

## Designs med ikke-randomiserede kontrolgrupper og pre-tests

- ▶ To grupper med flere pre-målinger

NR	$O_1$	$O_2$	X	$O_3$
NR	$O_1$	$O_2$		$O_3$

- ▶ To grupper med flere målinger og stimuli på forskellige tidspunkter

NR	$O_1$	X	$O_2$		$O_3$
NR	$O_1$		$O_2$	X	$O_3$

# Designs med ikke-randomiserede kontrolgrupper og pre-tests

- ▶ Modsat stimuli

NR	$O_1$	$X_+$	$O_2$
NR	$O_1$	$X_-$	$O_2$

- ▶ Beror på at skabe to stimulieffekter med modsat fortegn

# Designs med ikke-randomiserede kontrolgrupper og pre-tests

- ▶ Kohortekontrolgruppe

NR	$O_1$		
NR		X	$O_2$

- ▶ Kohortekontrolgruppe med pre-måling for begge grupper

NR	$O_1$	$O_2$			
NR			$O_3$	X	$O_4$

# Trusler til validiteten

- ▶ Talrige trusler til validiteten i alle stadier af den videnskabelige proces

## 1. I læsningen af litteraturen

- ▶ *Rhetoric biases, all's well literature bias, one-sided reference bias, positive result bias, hot stuff bias.*

## 2. I valg af sample

- ▶ *Popularity bias, centripetal bias, referral filter bias, diagnostic access bias diagnostic suspicion bias, unmasking bias, mimicry bias, previous opinion bias, wrong sample size bias, admission rate bias, prevalence-incidence bias, diagnostic vogue bias, diagnostic purity bias, procedure selection bias, missing clinical data bias, non-contemporaneous control bias, migrator bias, membership bias, non-respondent bias, volunteer bias.*

# Trusler til validiteten

## 3. I gennemførelsen af studiet

- ▶ *Contamination bias, withdrawal bias, compliance bias, therapeutic personality bias, bogus control bias.*

## 4. I målingen af variable

- ▶ *Insensitive measure bias, underlying cause bias, end-digit preference bias, apprehension bias, unacceptability bias, obsequiousness bias, expectation bias, substitution game, family information bias, exposure suspicion bias, recall bias, attention bias, instrument bias.*



# Trusler til validiteten

## 5. I analysen af data

- ▶ *Post-hoc significance bias, data dredging bias, scale degradation bias, tidying-up bias, repeated peeks bias.*

## 6. I fortolkningen af analysen

- ▶ *Mistaken identity bias, cognitive dissonance bias, magnitude bias, significance bias, correlation bias, under-exhaustion bias,*

## Få cases

- ▶ Hvad kan vi lære af få cases?
- ▶ Flere observationer er som regel bedre
- ▶ Ofte kan indsamling af flere data være vanskeligt
  - ▶ Der findes ikke flere observationer
  - ▶ Omkostningstungt

## Få cases: Casestudier

- ▶ Udfordringen er især intern validitet
- ▶ Fordrer ofte bestemte antagelser (Lieberson 1991):
  1. Deterministisk kausalitet
  2. Ingen målefejl
  3. Én årsag
  4. Ingen interaktionseffekter

# Få cases: Kan vi lære noget af $N=1$ ?

- ▶ Selvfølgelig
  - ▶ Samfundsvidenskaberne er fyldt med gode casestudier af én case
- ▶ Eksempler fra Science/Nature:
  - ▶ Kaminski et al. (2004): Word Learning in a Domestic Dog: Evidence for “Fast Mapping”
  - ▶ Ericsson et al. (1980): Acquisition of a memory skill
  - ▶ Anon (1970): Effects of Sexual Activity on Beard Growth in Man

## Få cases: Hvad med QCA?

- ▶ Qualitative comparative analysis (QCA)
- ▶ Kan fungere, men vær opmærksom på væsentlige metodiske udfordringer
  - ▶ Lucas og Szatrowski (2014): Qualitative Comparative Analysis in Critical Perspective
  - ▶ Krogslund et al. (2015): Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in fsQCA
  - ▶ Tanner (2014): QCA is of questionable value for policy research

# Få cases: Big data

- ▶ Observationsstudier er ikke kun relateret til mindre studier
- ▶ Big data/machine learning vinder frem i disse år
- ▶ Forudsigelser versus kausalitet
- ▶ Introduktionsartikler:
  - ▶ Breiman (2001): Statistical Modeling: The Two Cultures
  - ▶ Lazer et al. (2009): Computational Social Science
  - ▶ Varian (2014): Big Data: New Tricks for Econometrics

## Få cases: Matching

- ▶ Bliver vi altid klogere af at have flere data?
  - ▶ Med andre ord: Er flere data altid løsningen?
- ▶ Ikke alle observationer har en kontrafaktisk ækvivalent
- ▶ Dermed ikke alle observationer, der kan bruges i en kausalanalyse
  - ▶ Emnet for næste uge

# Afbrudt tidsseriedesign

- ▶ Det mest simple

---

$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$X$	$O_6$	$O_7$	$O_8$	$O_9$	$O_{10}$
-------	-------	-------	-------	-------	-----	-------	-------	-------	-------	----------

---

- ▶ Udvidet med endnu en gruppe

---

$O_1$	$O_2$	$O_3$	$O_4$	$O_5$	$X$	$O_6$	$O_7$	$O_8$	$O_9$	$O_{10}$
$O_1$	$O_2$	$O_3$	$O_4$	$O_5$		$O_6$	$O_7$	$O_8$	$O_9$	$O_{10}$

---



# Afbrudt tidsseriedesign

- ▶ To forskellige afhængige variable

---

$O_{A1}$	$O_{A2}$	$O_{A3}$	$O_{A4}$	$O_{A5}$	$X$	$O_{A6}$	$O_{A7}$	$O_{A8}$	$O_{A9}$	$O_{A10}$
$O_{B1}$	$O_{B2}$	$O_{B3}$	$O_{B4}$	$O_{B5}$		$O_{B6}$	$O_{B7}$	$O_{B8}$	$O_{B9}$	$O_{B10}$

---

- ▶ Tilføje og fjerne stimuli
  - ▶ Evt. en midlertidig politik

---

$O_1$	$O_2$	$O_3$	$X$	$O_4$	$O_5$	$O_6$	$X'$	$O_7$	$O_8$	$O_9$
-------	-------	-------	-----	-------	-------	-------	------	-------	-------	-------

---

# Afbrudt tidsseriedesign

- ▶ Udfordringer
  - ▶ Mange stimuli implementeres langsomt og diffust
  - ▶ Stimuli kan have tiltagende effekter og ikke være målbare med det samme
  - ▶ Mange tidsseriedata har ikke nok data til at lave statistiske analyser
    - ▶ Kræver mange observationer
  - ▶ Det kan være svært at få adgang til god data
  - ▶ Manglende data
  - ▶ Skiftende intervaller
  - ▶ Skiftende mål

# Afbrudt tidsseriedesign

- ▶ Eksempler med tværsnitsdata
- ▶ Udnytter information omkring, hvornår data er indsamlet
- ▶ Respondenter interviewet *før* en begivenhed fungerer som en approksimeret kontrafaktisk gruppe til dem, der er interviewet *efter* en begivenhed
- ▶ Overvej flere ting
  - ▶ Hvad er antagelserne?
  - ▶ Påvirker begivenheden sandsynligheden for at deltage i en undersøgelse?
  - ▶ Hvor stærkt er stimuli? (reagerer alle på begivenheden?)

# Eksempel: Effekten af terror på institutionel tillid (Dinesen og Jæger 2013)

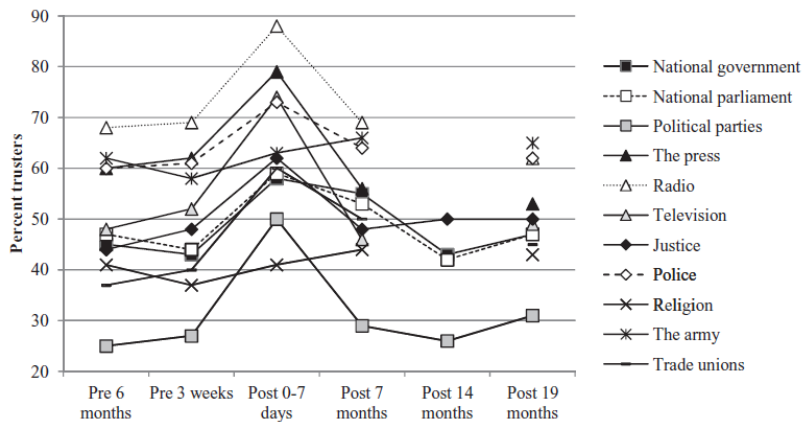


Figure 1. Trends in trust in institutions before/after 3/11 attack.

# Eksempel: Effekten af terror på holdningen til immigranter (Legewie 2013)

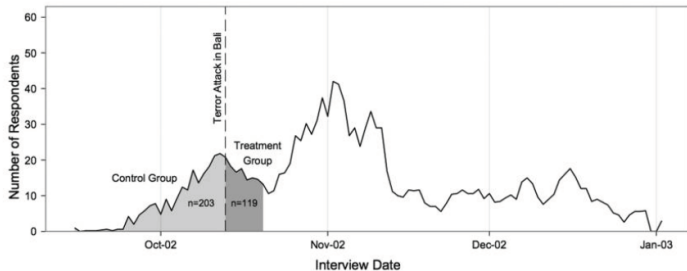


FIG. 1.—Treatment and control groups for Portugal. The figure shows the number of observations per day throughout the fieldwork period in terms of a four-day moving average. Control group (*light gray*) includes the respondents who were interviewed in the 30 days before the event. Treatment group (*dark gray*) includes the respondents who were interviewed in the week after the event.

# Eksempel: Effekten af terror på holdningen til immigranter (Legewie 2013)

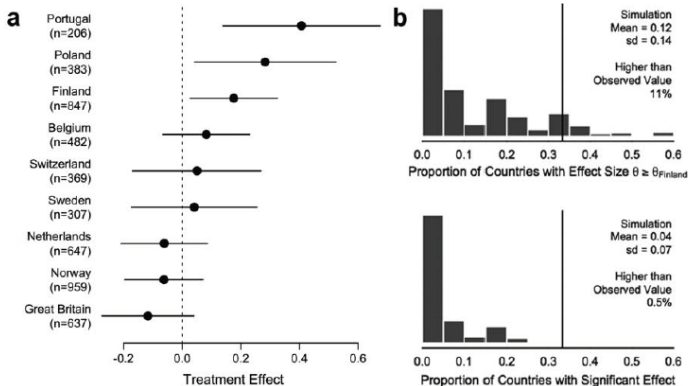
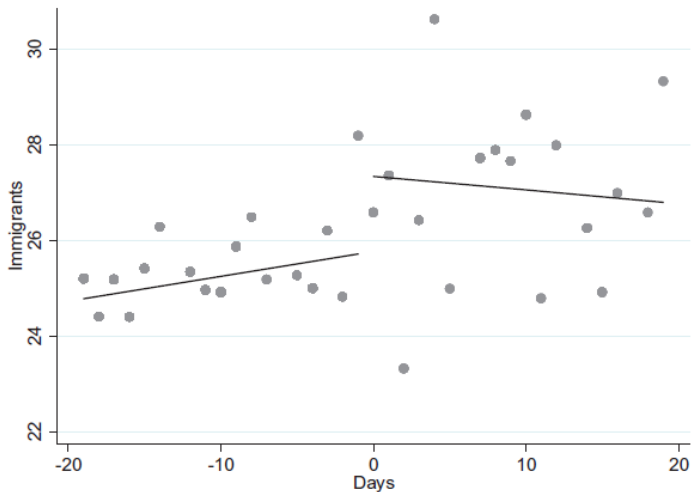


FIG. 3.—Estimated treatment effect for the terror attack in Bali. *a*, Estimated treatment effect; *b*, fictitious event simulation

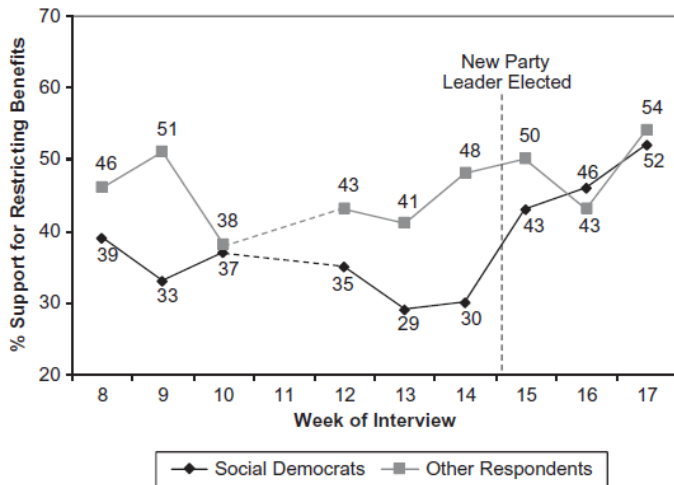
# Eksempel: Effekten af terror på holdningen til immigranter (Jakobsson og Blom 2014)

Figure 1

*Mean score on Immigrants by days, and regression lines before and after the attacks*



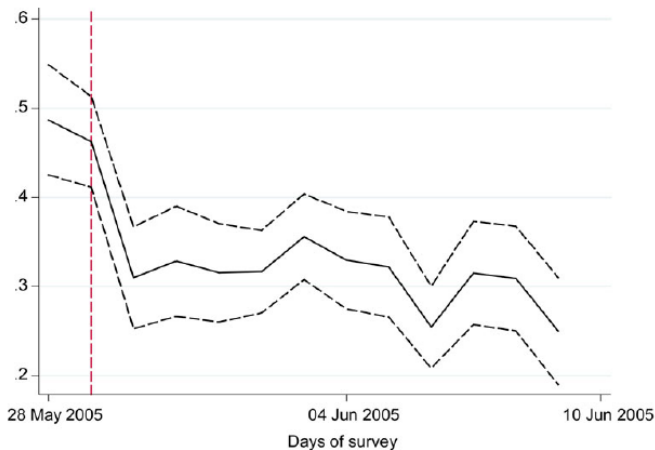
## Eksempel: Effekten af partiernes politik til holdningen til efterlønnen (Slothuus 2010)





# Eksempel: Effekten af økonomiske chock på opfattede levevilkår (Hariri et al. 2016)

FIGURE 1. Subjective Well-Being Around Time of Devaluation



Notes: Vertical line shows timing of devaluation. Dashed curves are 95% confidence intervals.

Source: Authors' analysis based on data sources discussed in text.

# Dagens øvelse

- ▶ Lav et diagram, der tester en kausal påstand
- ▶ Minimumskrav:
  - ▶ Der skal være et stimuli (uafhængig variabel)
  - ▶ Der skal være en postmåling (afhængig variabel)
- ▶ Diskuter de største validitetsudfordringer
  - ▶ Statistisk validitet
  - ▶ Intern validitet
  - ▶ Ekstern validitet
  - ▶ Konstruktionsvaliditet
- ▶ Diskuter hvordan designet bedst kan styrkes
  - ▶ Ideelt
  - ▶ Praktisk
  - ▶ Etisk?

# Næste gang

- ▶ Observationsdata, part II
- ▶ Torsdag 14. april, kl. 10-12 i U143