



UiO  **Institutt for spesialpedagogikk**
Det utdanningsvitenskapelige fakultet

Forskningsdesign: Eksperimentelle og kvasieksperimentelle. Indre validitet

Astrid Marie Jorde Sandsør



I dag

- Beskrivende forskning vs årsaksforskning
 - Årsaksforskning (Effektstudier)
 - Potensielle utfall
 - Indre validitet
 - Gode design for effektstudier
 - Eksperimenter
 - Kvasieksperimenter
- Slides fra SPED 4010,
(Zachrisson, 2020)

Beskrivelser vs. årsaker

Beskrivelser (hvordan)

- Kvalitativ eller kvantitativ
- Mening, sammenhenger, forskjeller, utvikling

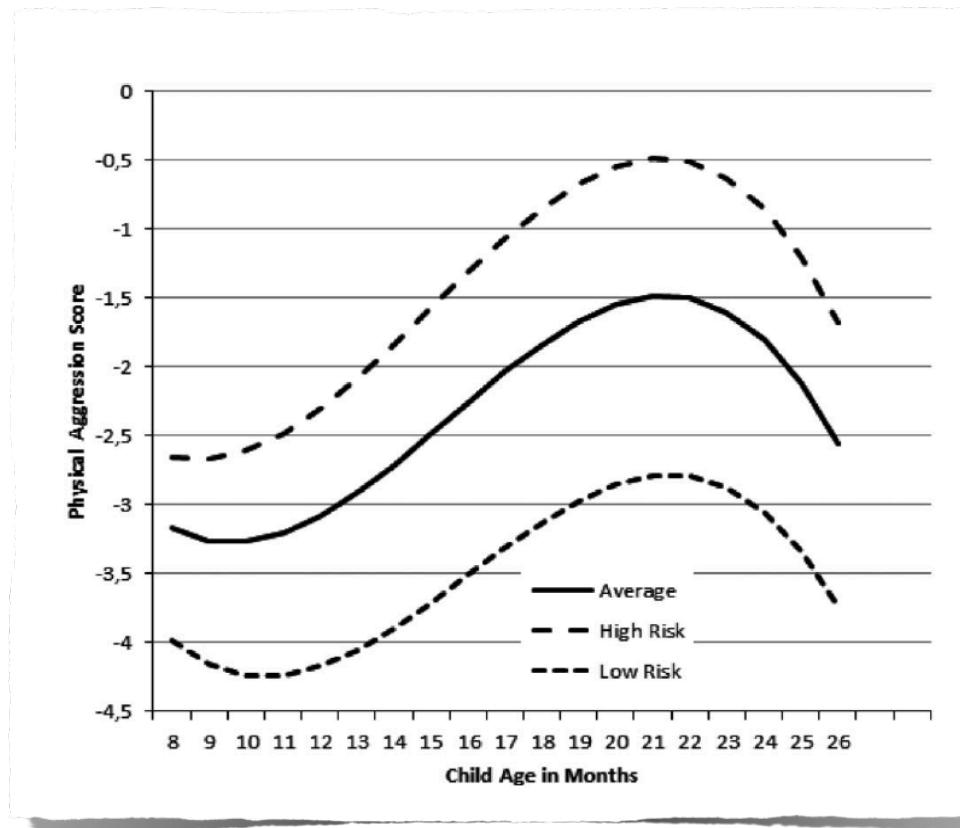
Årsaker (hvorfor)

- Noe som fører til noe annet
- «påvirker» «fører til»
«reduserer» «øker» «effekt»
«kausalitet»

Beskrivende forskning

Eksempel

- Hvordan utvikles aggressjon hos små barn?



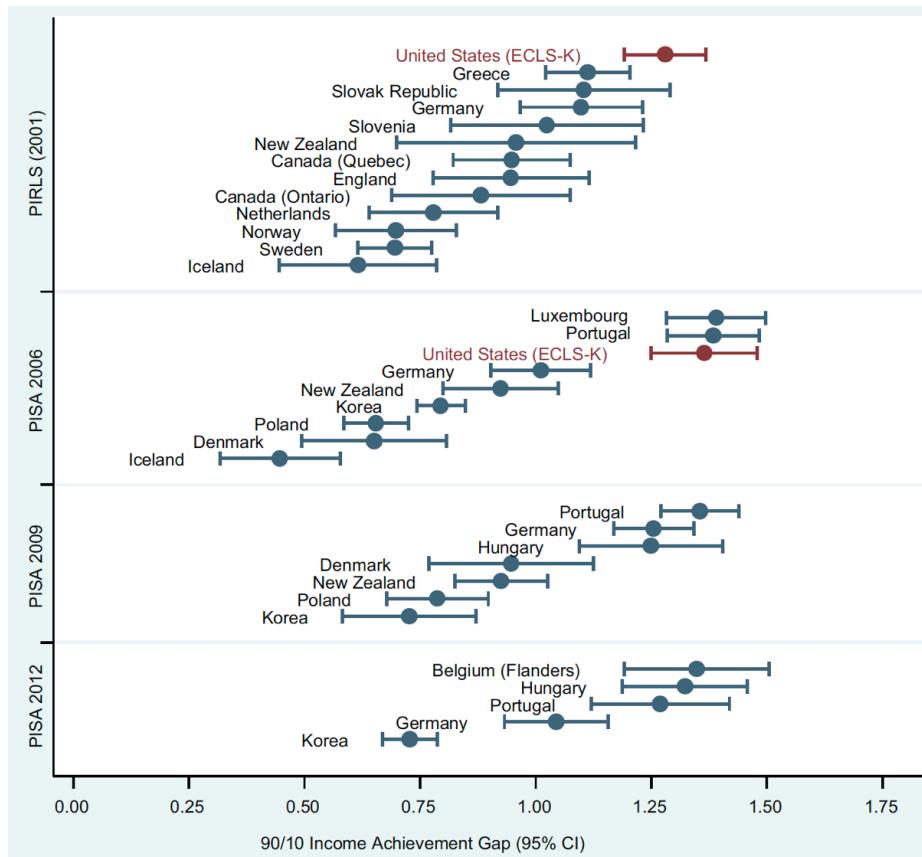
Beskrivende forskning

Eksempel

- Hvordan utvikles aggressjon (Avhengig variabel: Y) hos småbarn?
- Hva kjenneregner barn som utvikler mer aggressjon?
 - f.eks. kjønn, mors mentale helse, familieforhold? (Uavhengige variabler: X)
 - Data: Spørreskjema til mor, samlet over tid (longitudinell studie)

Beskrivende forskning

Eksempel



AERA Open
July-September 2016, Vol. 2, No. 3, pp. 1–2
DOI: 10.1177/233285841664951
© The Author(s) 2016. <http://ero.sagepub.com>

Patterns of Cross-National Variation in the Association Between Income and Academic Achievement

Anna K. Chmielewski
University of Toronto
Sean F. Reardon
Stanford University

- Hvor stor er forskjellen i testskårer (avhengig variabel: Y) mellom barn fra rike og fattige familier (uavhengig variabel: X) i ulike land?
 - Data: Testskårer (ILSA), foreldrerapport om inntekt

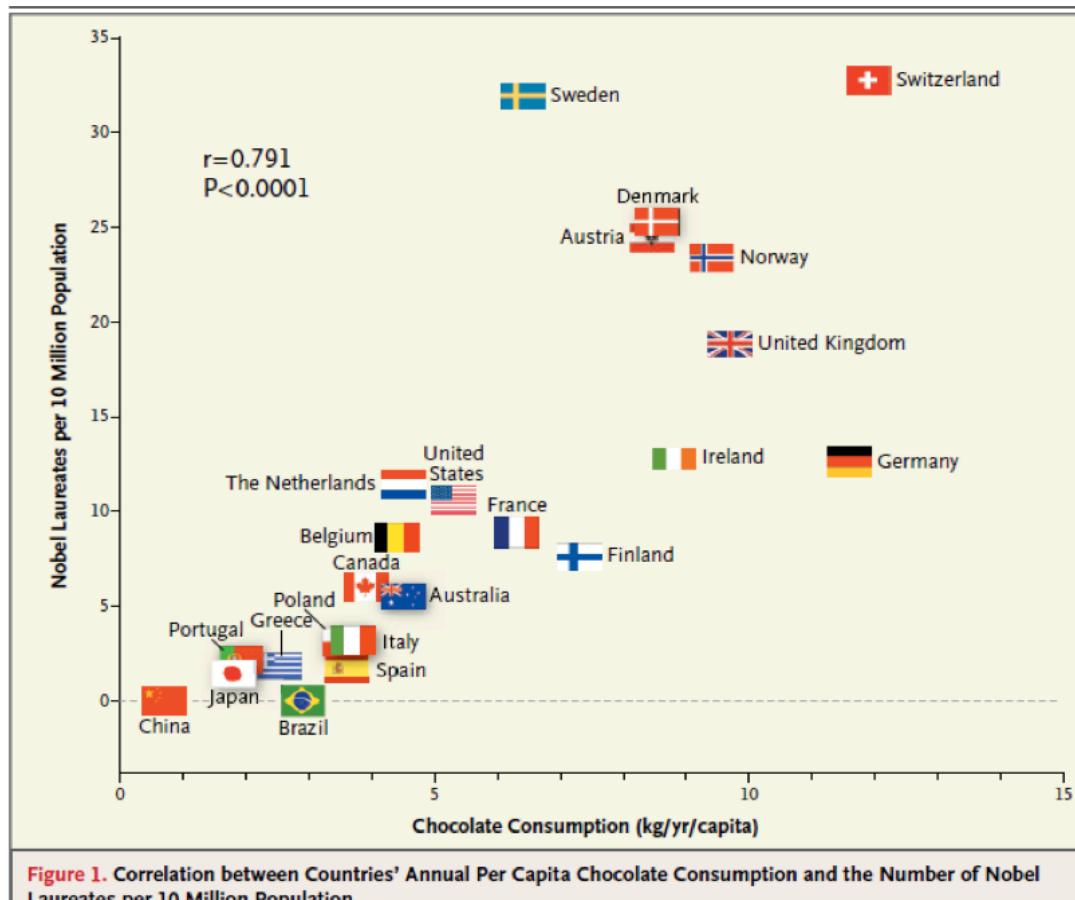
Beskrivende forskning

Hvordan? ←→ Hvorfor?

- Hva er problemstillingen?
- Er metoden egnet?
 - Kvaliteten på svaret avhenger av kvaliteten på studien
 - Validitet – er metoden valid for å gi et svar på spørsmålet?

Årsaksforskning / Effektstudier

Eksempel



The NEW ENGLAND JOURNAL of MEDICINE

OCCASIONAL NOTES

Chocolate Consumption, Cognitive Function, and Nobel Laureates

Franz H. Messerli, M.D.

Årsaksforskning / Effektstudier

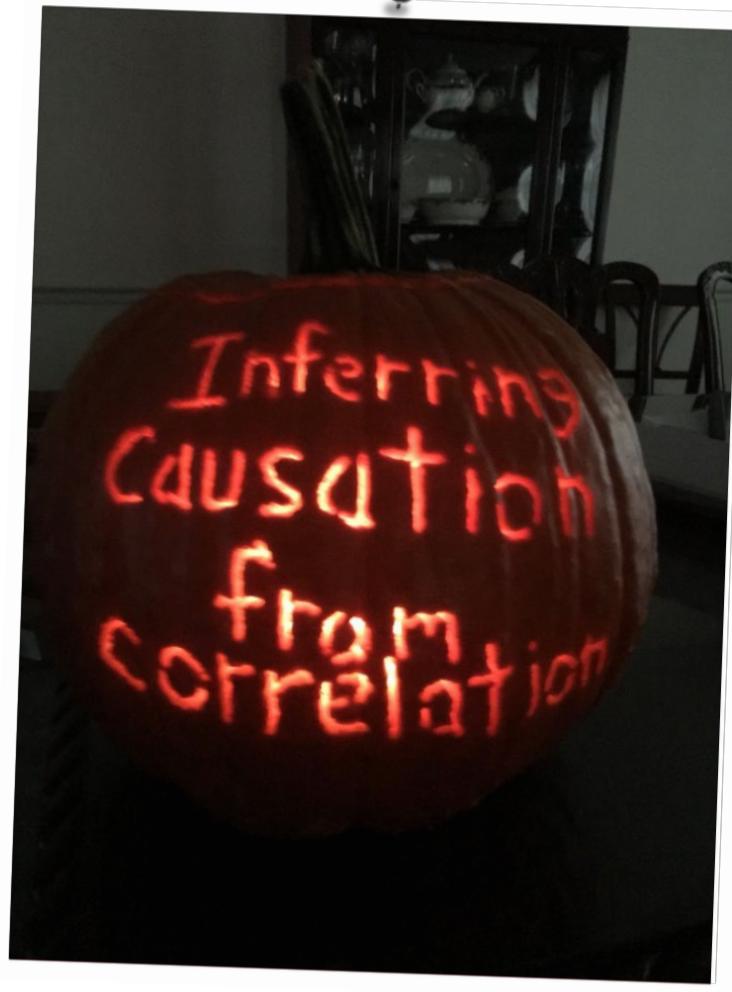
Eksempel

Since chocolate consumption could hypothetically improve cognitive function not only in individuals but also in whole populations, I wondered whether there would be a correlation between a country's level of chocolate consumption and its population's cognitive function. To my knowledge, no data on overall national cognitive function are publicly available. Conceivably, however, the total number of Nobel laureates per capita could serve as a surrogate end point reflecting the proportion with superior cognitive function and thereby give us some measure of the overall cognitive function of a given country.

<- Hypotese om årsaksforhold

<- Operasjonalisering av utfall (å gjøre et teoretisk begrep til en målbar størrelse)

Årsaksforskning / Effektstudier



Noe av det skumleste vi kan gjøre:

Å tolke årsak fra en sammenheng

Årsaksforskning / Effektstudier

Kausalitet: Litt historie

- Hume (1738-1740): «A Treatise of Human Nature»
 1. "The cause and effect must be contiguous in space and time.»
 2. "The cause must be prior to the effect.»
 3. "There must be a constant union betwixt the cause and effect. 'Tis chiefly this quality, that constitutes the relation.»
- Bollen's (1989) three components for causal relationship:
 1. Correlation / Association
 2. Temporality / Direction of influence
 3. Isolation (holde alt annet konstant bortsett fra årsaksforholdet)

Årsaksforskning / Effektstudier

To perspektiver

Rubin:

- Økonomi/Sosiologi/Medisin
- Potential outcomes/counterfactuals
- Formell definisjon & Heuristisk
- «hva hadde skjedd hvis»

Campbell (Shadish, Cook)

- Psykologi/Utdanning
- Validitet (indre)
- Trusler mot validitet
- Design

Årsaksforskning / Effektstudier

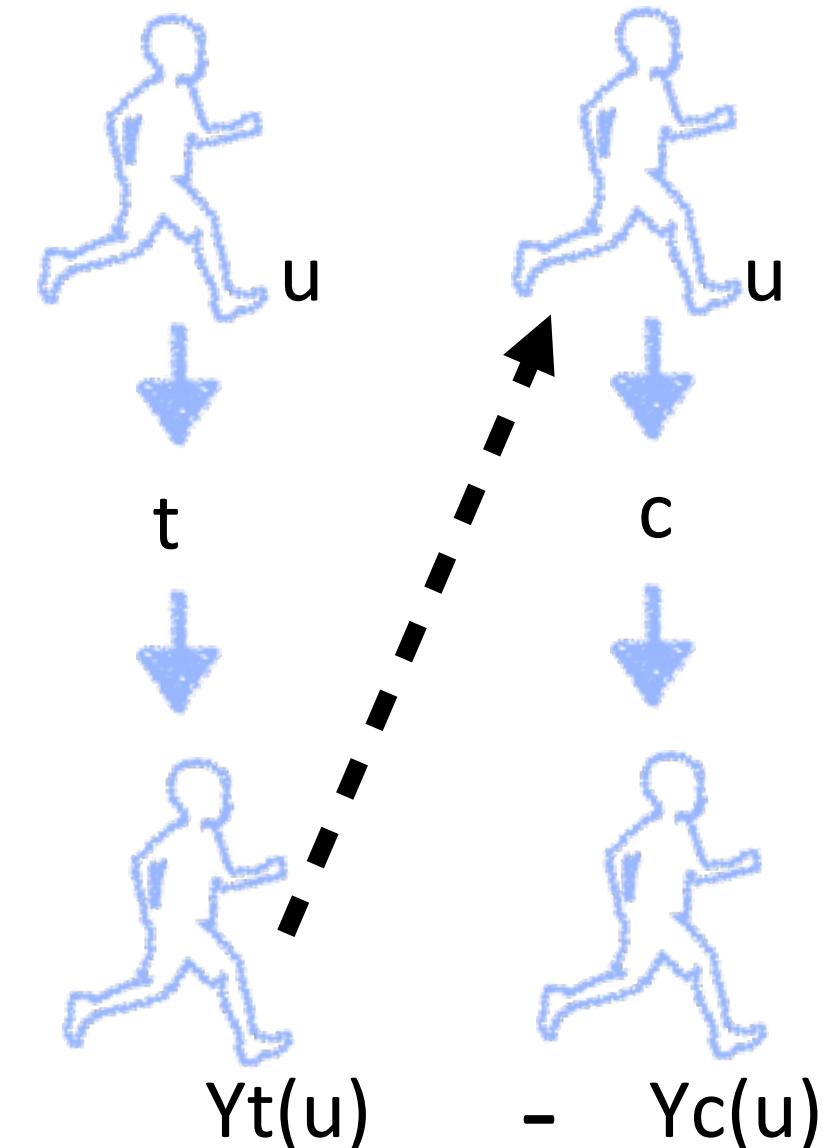
Er det kausalt?

- Er X (f.eks. tid med spesialpedagog) assosiert/korrelert med Y (bedre skoleresultater)?
- Endrer Y seg om vi endrer på X?

Potential outcomes

Rubin

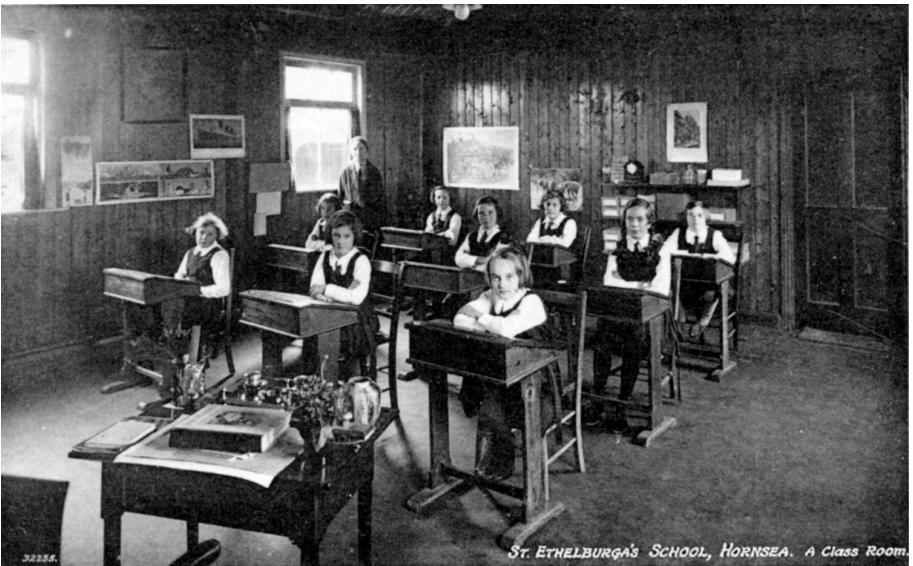
1. Gi barn u tiltak (t), mål utfallet ($Y_t(u)$).
2. Gi barn u kontroll-betingelsen (c), på samme tidspunkt og sted, mål utfallet ($Y_c(u)$).
3. Kausal effekt = $Y_t(u) - Y_c(u)$



Potential outcomes

- «Hva hadde skjedd hvis ikke»
- Det kontrafaktiske – det motsatte av det som faktisk skjedde
 - Det faktiske: Sveitsere spiser mye sjokolade og får mange nobelpriser
 - Det kontrafaktiske: Sveitsere spiser ikke mye sjokolade og...?
- Tankemodell: Hva vi mener med årsak
- Spør alltid: Hva hadde skjedd hvis «årsaken» ikke hadde vært der

Det perfekte eksperiment



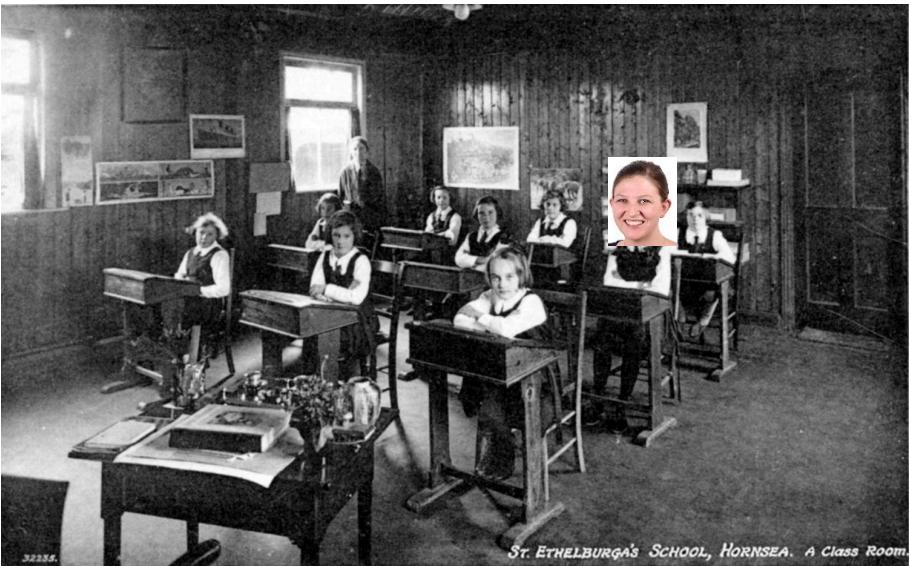
Bradford Timeline, Flickr; nifu.no

Hva er effekten på
testresultater av at elever
plasseres i mindre klasser?

Fører mindre klasser til bedre testresultater?



Det perfekte eksperiment



Bradford Timeline, Flickr; nifu.no

Hva er effekten på
testresultater av at elever
plasseres i mindre klasser?

Fører mindre klasser til bedre testresultater?



Det perfekte eksperiment



Bradford Timeline, Flickr; nifu.no

Vi kan aldri observer samme person samtidig i både små og store klasser:
Vi kan aldri observere det kontrafaktiske

Fører mindre klasser til bedre testresultater?



Boston Public Library , Flickr; nifu.no

Potential outcomes

En kasual slutning er som en simulert tidsmaskin

Kausalitet kan oppnås med:
eksperiment, kvasi-eksperiment,
statistisk kontroll



Validitet

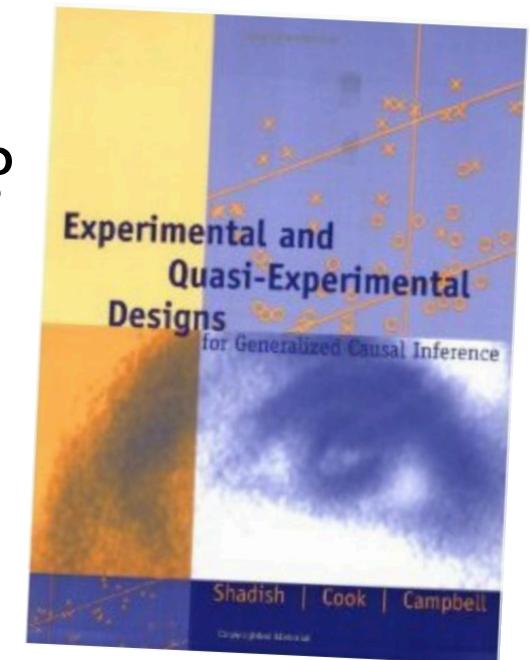
- «Validity refers to the approximate truth of an inference»
(Shadish et al)

Shadish, Cook, & Campell
System av validitet

Den omtrentlige sannheten ved en slutning (konklusjon)

- Tror vi at det denne studien viser er sant?

Tror vi at man blir smartere av å spise mer sjokolade?



Validitet

Hovedtyper validitet:

- Indre validitet: Årsak
- Statistisk validitet: Er statistikken god?
- Begrepsvaliditet: Måler vi det vi tror vi måler på en god måte?
- Ytre validitet: Generalisering

Trusler mot valide slutninger:

- Mulige (possible) grunner til feilslutning
- Sannsynlige (probable) grunner til feilslutning

Indre validitet

- Viser studien et årsaksforhold?
 - Er sjokolade årsaken til Nobelpriser?
- Trusler: Er det alternative, plausible, forklaringer?
 - NB! Ikke alle trusler er plausible i alle studier
- Tenke gjennom ved design av en studie:
 - Hvordan kan vi gardere oss mot alternative forklaringer?

Trusler mot indre validitet

- Ikke-observert konfundering
- Retning
- Historie
- Modning
- Seleksjon
- Testing
- Frafall
- Instrumentering
- Regresjon mot gjennomsnittet
- Atypisk kontrollgruppe

Trusler mot indre validitet

Ikke-observert konfundering:

- Ved statistisk kontroll
 - Er alle plausible alternativer inkludert i modellen
- Ved design-basert slutning
 - Er alternative forklaringer ekskludert gjennom designet?

Retning:

- Er uavhengig variabel (X) årsak til avhengig variabel (Y), eller motsatt?
 - (er sjokolade årsak til nobelpriser, eller nobelpriser årsak til sjokolade?)
- Tiltak: Sikre at X kommer før Y (ikke bare er målt før)

Trusler mot indre validitet

Historie:

- Er det hendelser som oppstår uavhengig av X, og som forklarer Y
 - eks. Vi sammenligner to grupper før og etter en reform eller endring i kultur
- Tiltak: Sikre at vi vet at gruppene kommer fra samme tid/sted

Modning:

- Endringer som skyldes naturlig vekst/modning/endring
 - eks. Endrer grupper seg ulikt (språk hos 3-4 vs. 5-6-åringer)
- Tiltak: Sikre at vi har (plausibel) grunn til å tro at vi sammenligner personer i samme modningsprosess

Trusler mot indre validitet

Seleksjon:

- Sammenligner vi epler og pærer?
 - Selv-seleksjon inn i tiltak?
 - eks. er barn som begynner i bhg ved 12 mnd ellers like de som begynner ved 24 mnd
- Tiltak: Randomisering, plausibel kvasi-randomisering, perfekt statistisk kontroll

Testing:

- Blir noen testet flere ganger enn andre?
 - Eks. lærer av testen
- Tiltak: Sikre at alle testes likt

Trusler mot indre validitet

Regresjon til gjennomsnittet:

- «Ekstreme» skårer varer sjeldent over tid (uteliggere)
- Problem hvis vi selekterer på Y (velger f.eks. barn med høy/lav skåre til et tiltak)
- Tiltak: Ha kontrollgruppe, evt. mange repeterte målinger

Atypisk kontrollgruppe:

- Kontrollgruppen «saboterer» ved å søke tiltak
 - eks. skole som ikke får et tiltak prøver selv å sette i gang tiltaket
- Tiltak: Monitorer hva kontrollgruppe gjør

Ta-med-hjem

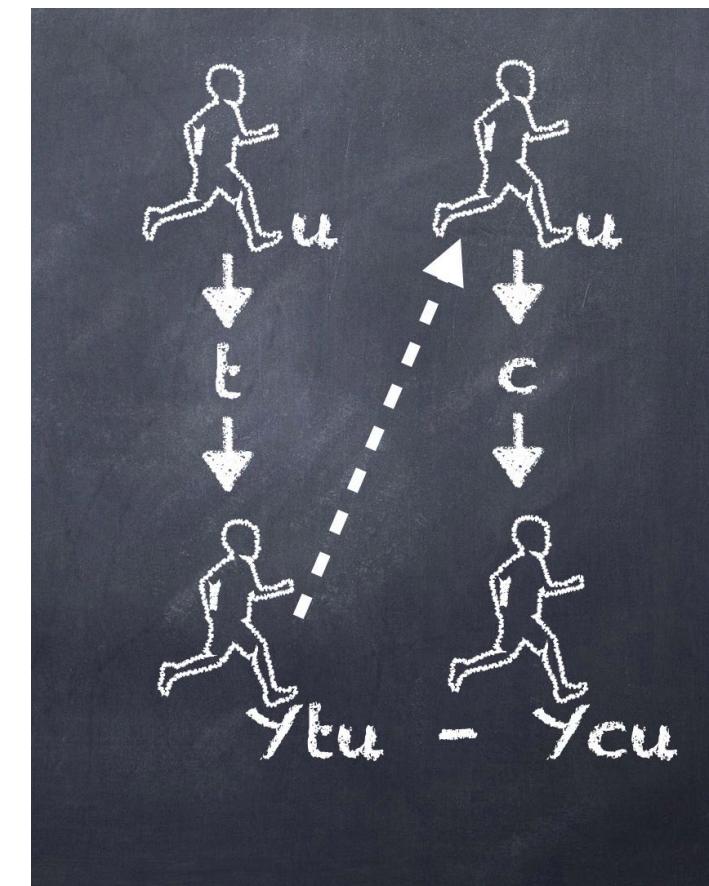
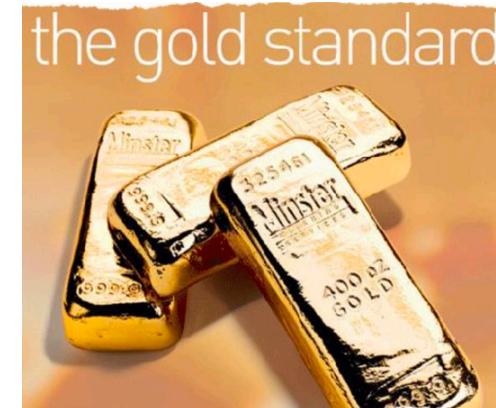
- Kausale slutninger
- Endring i X fører til endring i Y
- Potensielle utfall («hva hvis?»)
- Trusler mot indre validitet
- Tiltak

Gode design for effektstudier

- Hvordan studien er laget, og hva det betyr for indre/ytre validitet
- Kontrollere for trusler mot indre/ytre validitet
- Kontroll gjennom design
 - Eksperiment
 - Kvasi-eksperiment
- Kontroll gjennom statistikk
 - Ikke-eksperiment

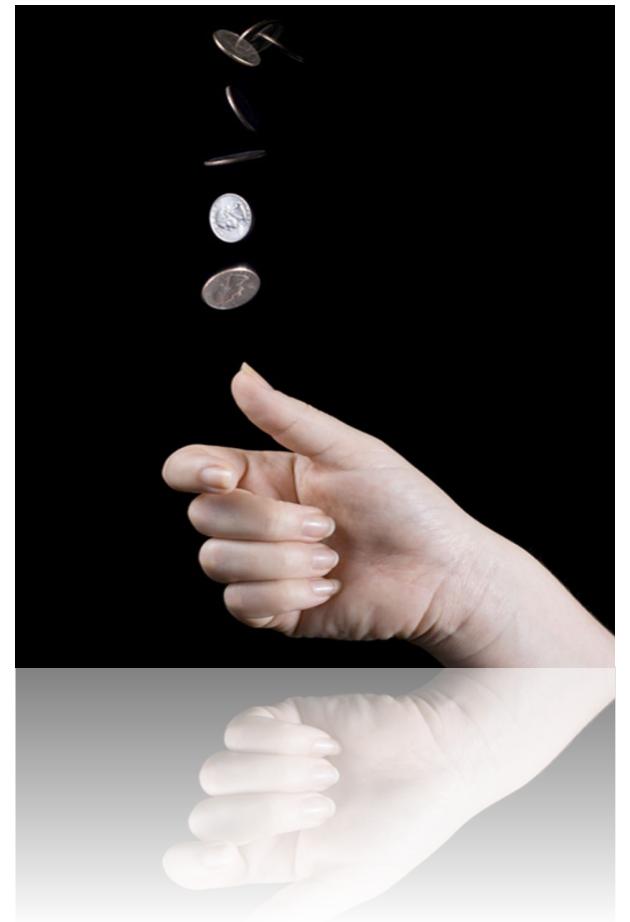
Eksperiment

- Design for sterk indre validitet
 - Teste effekter, f.eks. av tiltak
- Kalles også RCT
 - Randomized Controlled Trial



Eksperiment

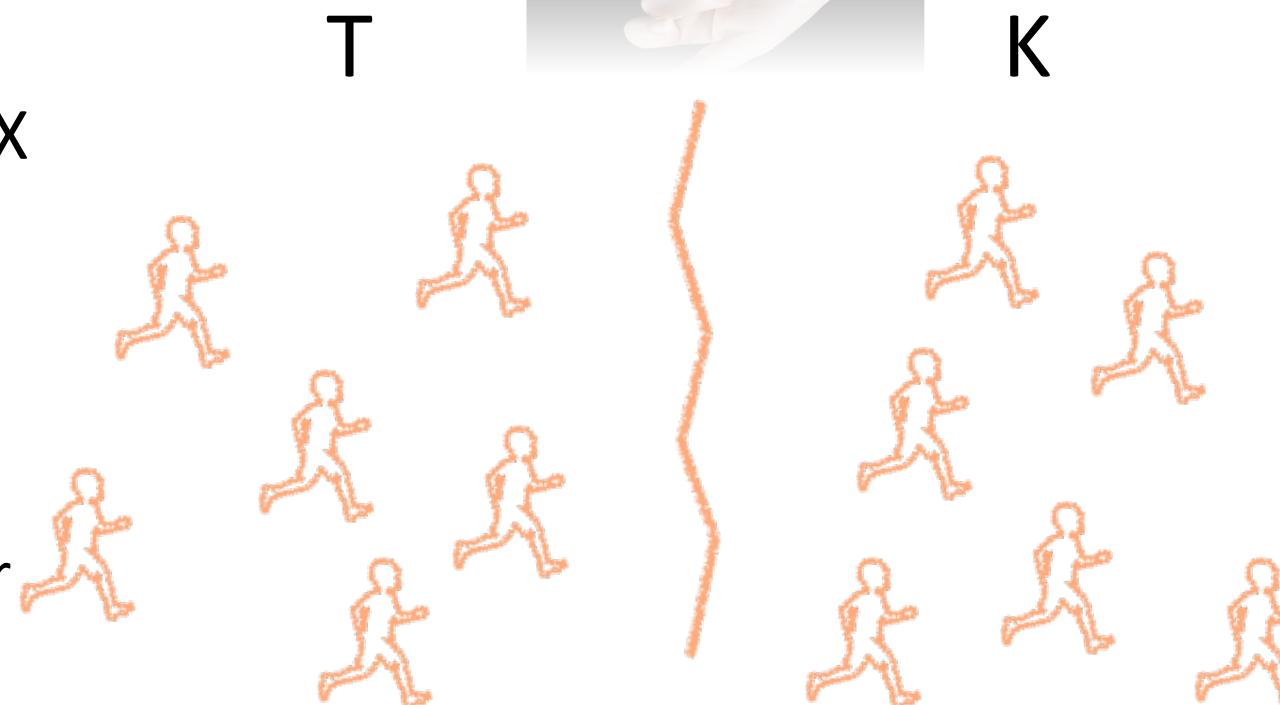
- Gruppe barn/klasser/skoler
- Tilfeldig fordeling
 - Tiltak (T) (tiltaksgruppe / forsøksgruppe)
 - Kontroll (K) (kontrollgruppe / sammenligningsgruppe)
- Gruppene er like bortsett fra T/K
- Gruppforskjellene i Y (utfall etter tiltak) skyldes T/K



Eksperiment

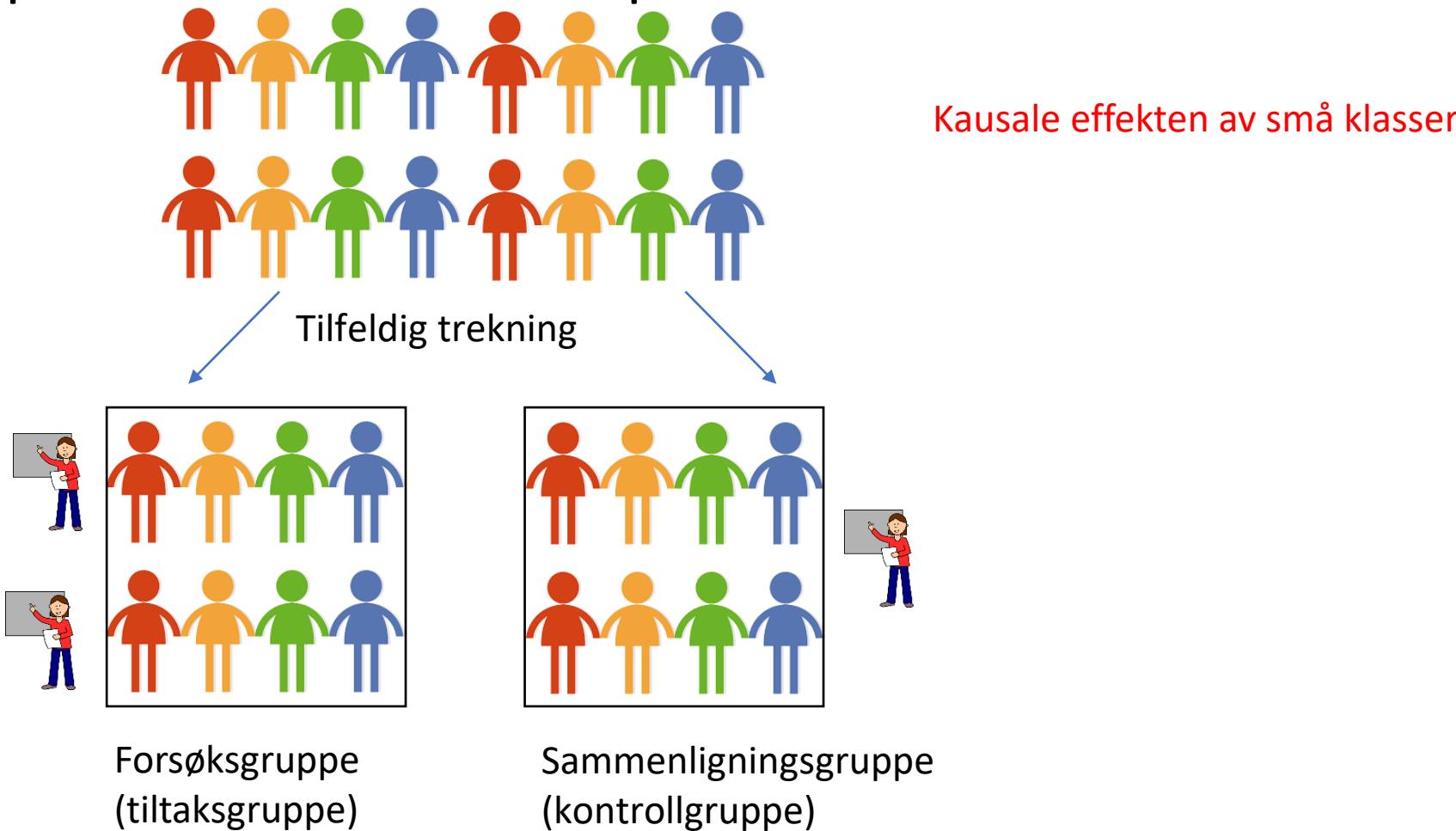


- Vi har kontroll på X (uavhengig variabel)
- Noen får - andre får ikke
 - eneste forskjell er X (T/K)

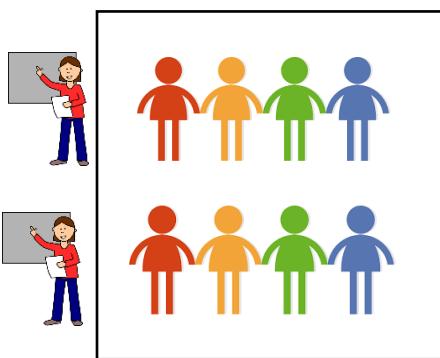


- Gjør gruppene (sannsynlig) like
- Utelukker alternative forklaringer
- Utelukker andre trusler (modning, seleksjon)

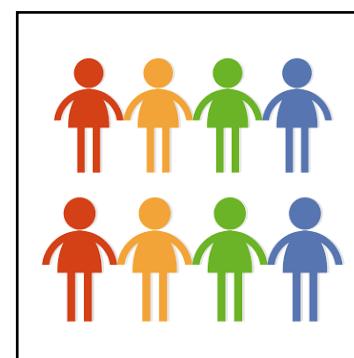
Eksperiment: Eksempel



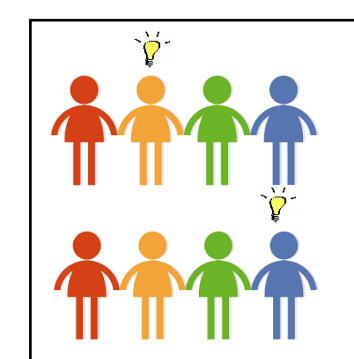
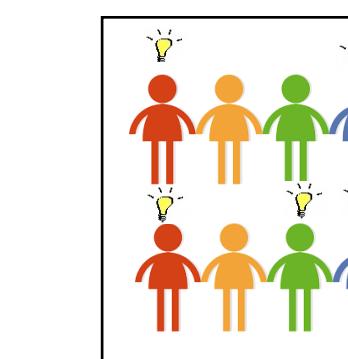
Pixabay.com



Forsøksgruppe



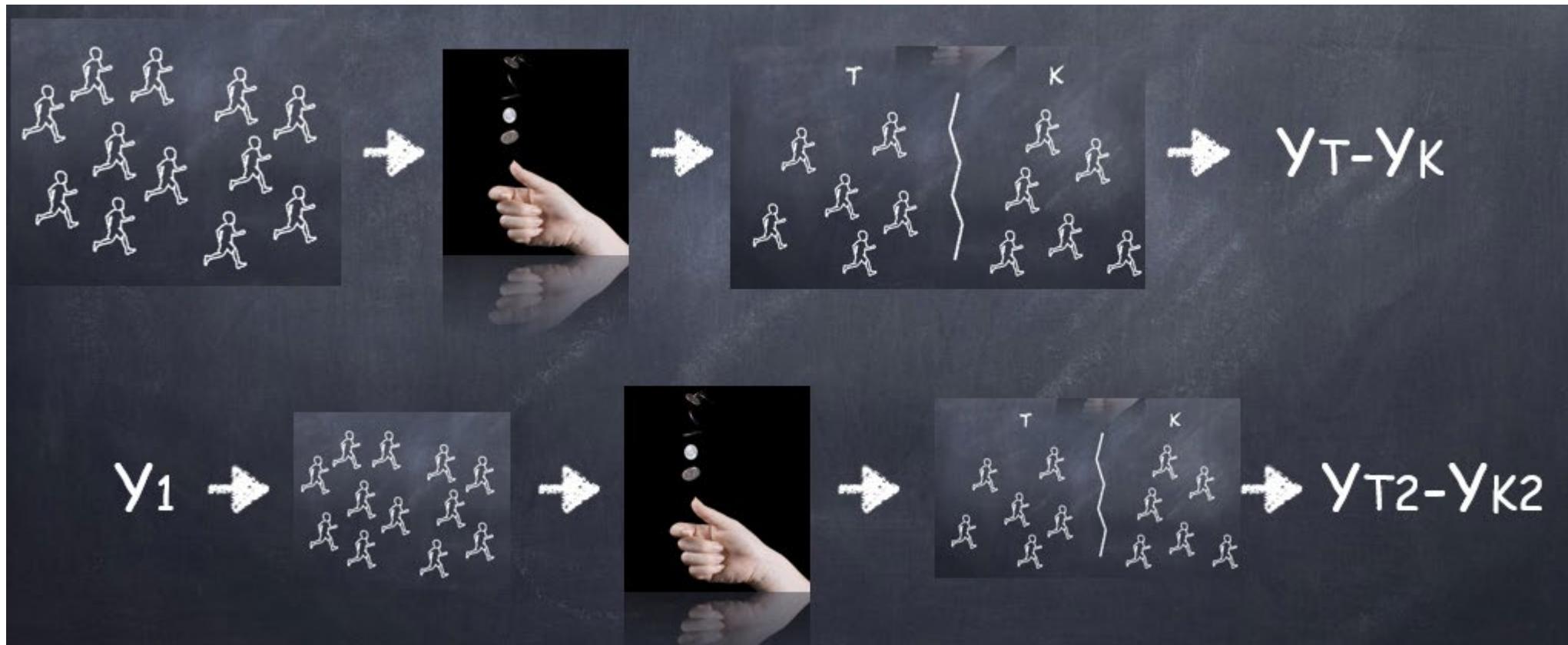
Sammenligningsgruppe



Kausale effekten av små klasser

Pixabay.com

Typer eksperiment -uten eller med premål



Eksperiment

- Hva kan vi teste?
 - Ting som kan manipuleres (påvirkes)
 - F.eks. trening/behandling/policy (f.eks. kontantstøtte, klassestørrelse)
- Hva kan vi ikke teste?
 - Ting som ikke kan manipuleres
 - Praktisk umulig, f.eks. kjønn
 - Etisk umulig, f.eks. amming

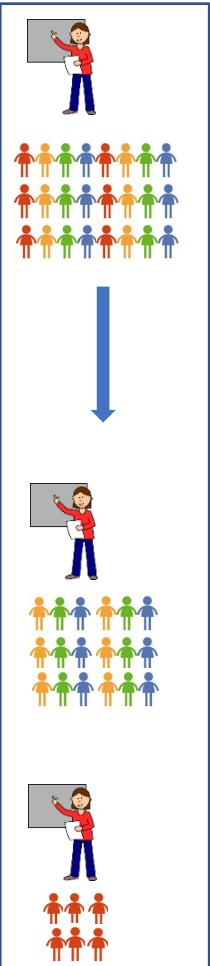
Eksperiment

- Hvem som får tilgang til tiltaket velges ved tilfeldig trekning
- Sikrer at det ikke er systematiske forskjeller mellom de som får tiltaket (forsøksgruppen) og de som ikke får det (sammenligningsgruppen)
- Observerbare forskjeller (f.eks. familiebakgrunn) er balansert i begge grupper
- Uobserverbare forskjeller (f.eks. motivasjon, talent) er også balansert i begge grupper
- I snitt vil forsøksgruppen og sammenligningsgruppen være sammenlignbare når forsøket starter og ville ha fulgt samme utvikling i fravær av tiltaket
- En sammenligning av gruppene i etterkant av tiltaket gir oss den kausale effekten av tiltaket

- Hvem-hva-hvor:
- 1+1: Et forskningsprosjekt ledet av NIFU, SØF og ISF.
- Lærer elevene mer matematikk gjennom å delta i smågruppeundervisning?
- En randomisert kontrollert studie
- Asker, Bodø, Bodø, Bodø, Bodø, Bodø, Bodø, Bodø, Trondheim, Tromsø, Ålesund
- 160 skoler deltar i prosjektet, tilfeldig fordelt i to grupper:
 - Forsøksgruppe (80 skoler): Får et ekstra lærerårsverk per skole i fire år
 - Sammenligningsgruppe (80 skoler): Fortsetter som før
 - Tidsramme: Fire skoleår, fra høsten 2016 til våren 2020



Om prosjektet



- En ekstra lærer brukes til å gi smågruppeundervisning i matematikk

- Tiltak for alle elever, både høyt-, middels- og lavtpresterende
- Pull-out strategi:
 - Grupperstørrelse 4-6 elever
 - Minst treskoletimer i uka
 - Varighet: 4-6 uker
 - To perioder hvert skoleår
 - Prosjektet anbefaler relativt homogene grupper

Skoleår: Kohort:	2016/17	2017/18	2018/19	2019/20
2008	3. klasse	4. klasse		
2009	2. klasse	3. klasse	4. klasse	
2010				4. klasse
2011			2. klasse	3. klasse

- Måling av elevenes ferdigheter i regning før forsøket begynner (pretest)
- Måling av elevenes ferdigheter i regning ved utgangen av skoleåret (posttest)
- Registreringsskjema
- Spørreundersøkelse til lærere og skoleleder
- Besøk ved skolene (casestudier)
- Registerdata (kjennetegn ved elev/foreldre og nasjonale prøver)
- Innhenting av samtykkeskjema for kobling av data

Eksperiment - begrensninger

- Generell likevekt
 - Hva er effekten av små klasser – når vi har tilgjengelige lærere
 - Hva er effekten når det er få eller mindre kvalifiserte lærere?
- Hawthorne effekter
 - Forsøksgruppen påvirkes av at de er med i et forsøk
- John Henry effekter
 - Sammenligningsgruppen påvirkes av at de er med i et forsøk
- Kan forsøket replikeres?
- Blir resultatene de samme med en annen populasjon?
- Blir resultatene de samme hvis tiltaket utformes litt forskjellig?

Eksperiment – trusler og fallgruver

- Brudd på randomisering
- Glemme spillover-effekter
- Ikke ha nok styrke
- Ikke ha gode nok utfallsmål
- Datainnsamling forskjellig i forsøks- og sammenligningsgruppe
- Høyt frafall i datainnsamling
- Glemme å undersøke implementering og andre trussler underveis
- Ytre validitet:
 - Er det som skjer i eksperimentet gyldig uten eksperimentet?
 - Ser det likt ut i verden?

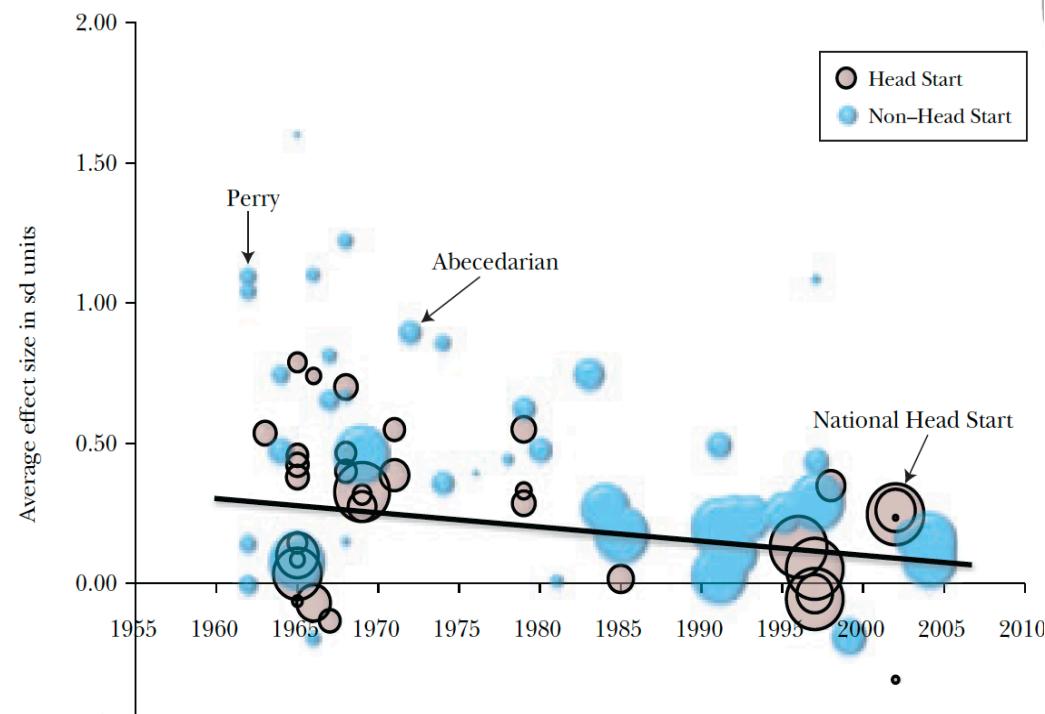
Investing in Preschool Programs[†]

Greg J. Duncan and Katherine Magnuson

Journal of Economic Perspectives—Volume 27, Number 2—Spring 2013—Pages 109–132

Figure 2

Average Impact of Early Child Care Programs at End of Treatment
(standard deviation units)



Improving the General Language Skills of Second-Language Learners in Kindergarten: A Randomized Controlled Trial

Kristin Rogde^a, Monica Melby-Lervåg^a, and Arne Lervåg^a

ABSTRACT

Second-language learners display poorer general language skills in the language used at school than their monolingual peers, which is a concern because general language skills (vocabulary, grammar, language expression, and comprehension) provide the foundation for later academic success. In a randomized controlled trial, we examined the efficacy of an intervention to improve second-language learners' general language skills in their second language. One hundred and fifteen children (mean age = 5.5 years) were randomly assigned to a business-as-usual control group or to an 18-week intervention program conducted by kindergarten teachers during the last semester of kindergarten. The children were assessed at pre-intervention, post-intervention and a seven-month follow-up. The children in the treatment group demonstrated significant improvements on a custom measure of taught vocabulary. The language intervention program also produced effects that generalized to standardized measures reflecting expressive language skills. These findings suggest that intervention programs designed to enhance second-language learners' expressive language in the second language can be successfully implemented in kindergarten settings before school entry.

KEYWORDS

second-language learners
general language skills
vocabulary
intervention
randomized controlled trial

Ta-med-hjem

- Eksperiment/RCT: Gullstandarden for kausale slutninger
- Tilfeldig fordeling T/K
- Primær trussel: ytre validitet

Kvasi-eksperiment

- «An experiment where units are not assigned to conditions randomly» (Shadish et al., 2002)
 - Barn «velges» til tiltak og kontroll av en grunn
 - Brudd på randomisering
 - MEN... kan allikevel være mulig å utnytte tilfeldighet for å finne kausalitet
- Fordeler
 - Kan ha høy ytre validitet
 - Testet i «den virkelige verden»
 - Kan teste hypoteser som ikke kan testes eksperimentelt
- Ulemper
 - Må sannsynliggjøre intern validitet

Kvasi-eksperiment

To tradisjoner:

1. «Sloppy» design som ikke sannsynliggjør indre validitet
 - Noe tradisjonell evalueringsforskning
 - Et dårlig eksperiment...
2. Gode design som sannsynliggjør indre validitet
 - «Nothing is wrong with making assumptions (...). The quality of these assumptions, not their existence, is the issue» (Little & Rubin, 2000)

Kvasi-eksperiment

The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2021



III. Niklas Elmehed © Nobel Prize Outreach.

David Card

Prize share: 1/2



III. Niklas Elmehed © Nobel Prize Outreach.

Joshua D. Angrist

Prize share: 1/4



III. Niklas Elmehed © Nobel Prize Outreach.

Guido W. Imbens

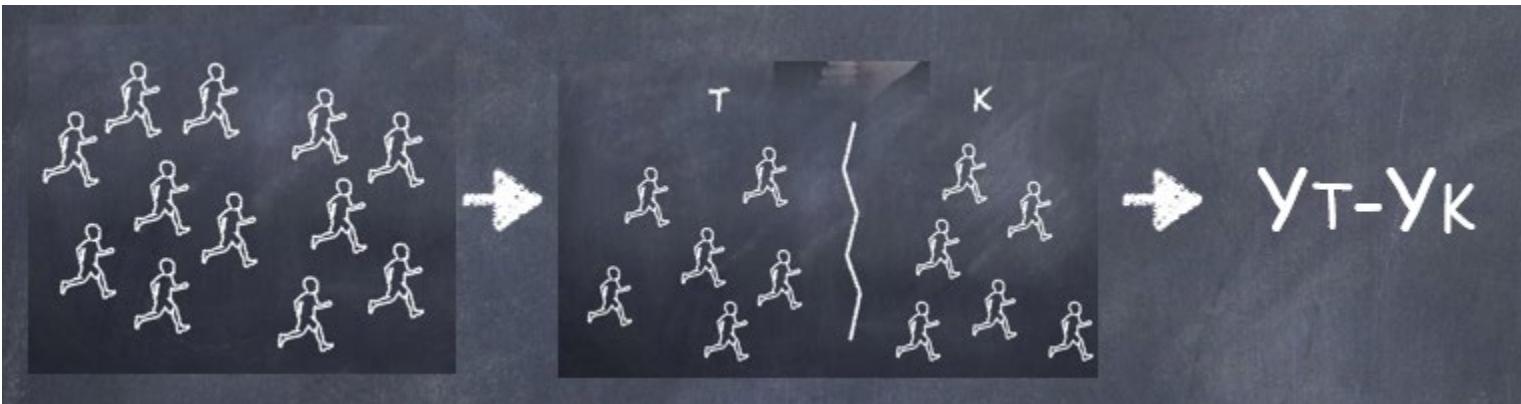
Prize share: 1/4

Natural experiments help answer important questions for society

This year's Laureates – David Card, Joshua Angrist and Guido Imbens – have provided us with new insights about the labour market and shown what conclusions about cause and effect can be drawn from natural experiments. Their approach has spread to other fields and revolutionised empirical research.

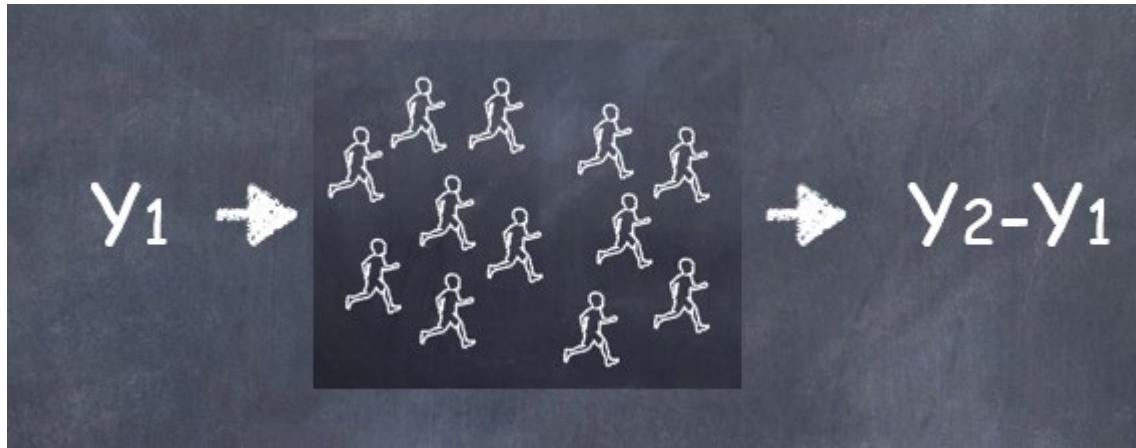
«Kredibilitetsrevolusjonen»

Kvasi-eksperiment – post-test design med ikke ekvivalente grupper («Sloppy»)



- Primær trussel er seleksjon
- Gruppene er ikke like:
 - Forskjell i Y kan skyldes intervension
 - Eller forskjeller i hvem som «velger» intervension

Kvasi-eksperiment – pre-post-test design uten kontrollgruppe («Sloppy»)



- Primær trussel er seleksjon og modning/historie
- Deltagere kan forandre seg av mange grunner
 - Forskjell i Y kan skyldes modning, regresjon til gjennomsnittet
 - Ukjent kontrafaktisk betingelse

Kvasi-eksperiment – kan vi gjøre det bedre?

- Avbrutte tidsserier (Interrupted time series, ITS) (litt «Sloppy»)
- Forskjell-i-forskjeller (Difference-in-differences, DD)
- Instrument variabler (Instrumental variables, IV)
- Regresjonsdiskontinuitet (Regression discontinuity, RD)
- Eksplisitte antagelser om kontrafaktiske betingelser og tester som sannsynliggjør disse

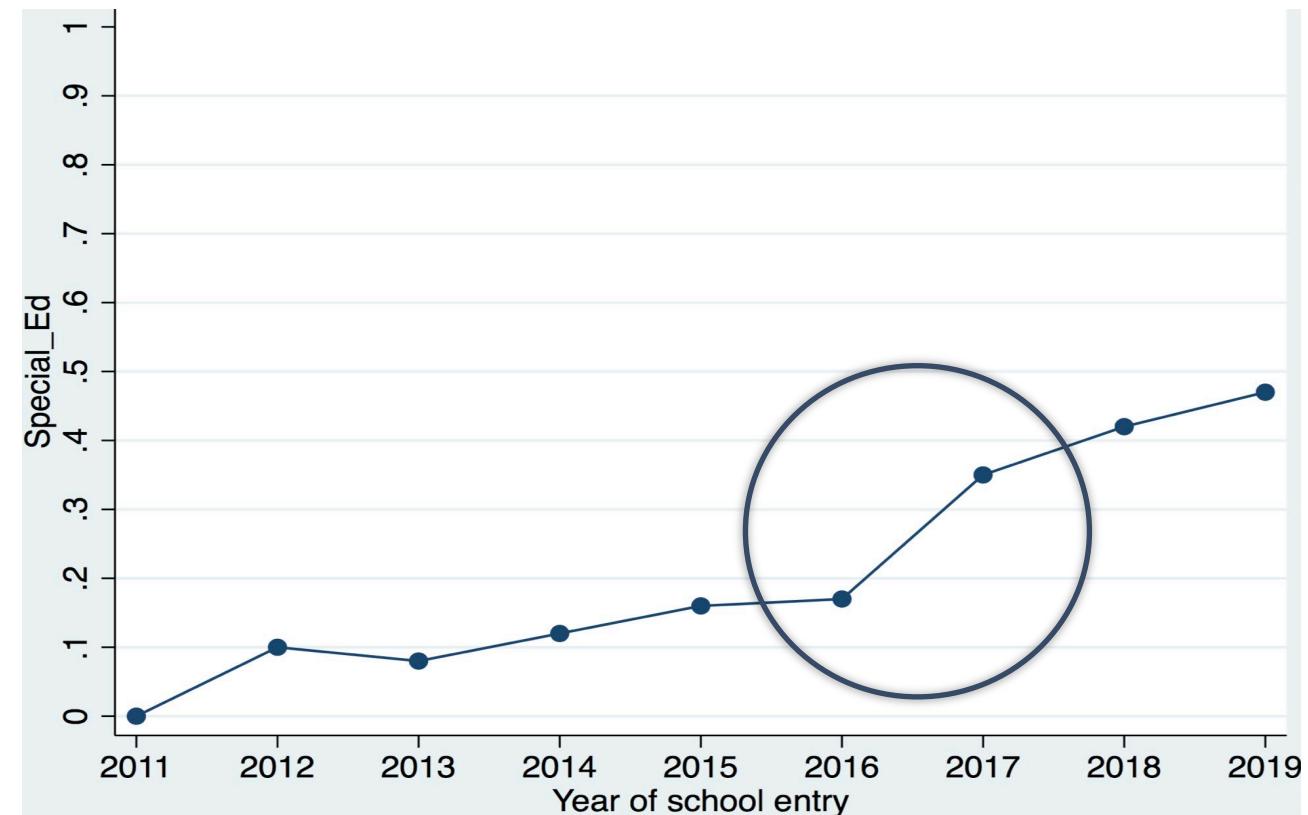
Kvasi-eksperiment – eksempel

- Reglene om rett til spesialpedagogisk hjelp og rett til tegnspråkopplæring for barn under opplæringspliktig alder ble flyttet fra opplæringsloven til barnehageloven 1. august 2016
- Gjør elever som får vedtak om spesialundervisning etter 1. august 2016 det bedre på kartleggingsprøver i 1. klasse, sammenlignet med elever som får vedtak om spesialundervisning før 1. august 2016?



Kvasi-eksperimenter: Avbrutt tidsserie

- Forskjellen i resultat på kartleggingsprøve i 1. klasse for de med vedtak om spesialundervisning før og etter tiltak



Men... hva om noe annet skjedde samtidig?

- Det var mer fokus på språkutvikling i barnehagen
 - F.eks. SPROFF (våren 2016)
- Mer videreutdanning av barnehageansatte
- PPT jobbet tettere opp mot barnehagene etter endringen

Kvasi-eksperimenter: Forskjell-i-forskjeller

Tiltaksgruppe vs.
sammenligningsgruppe

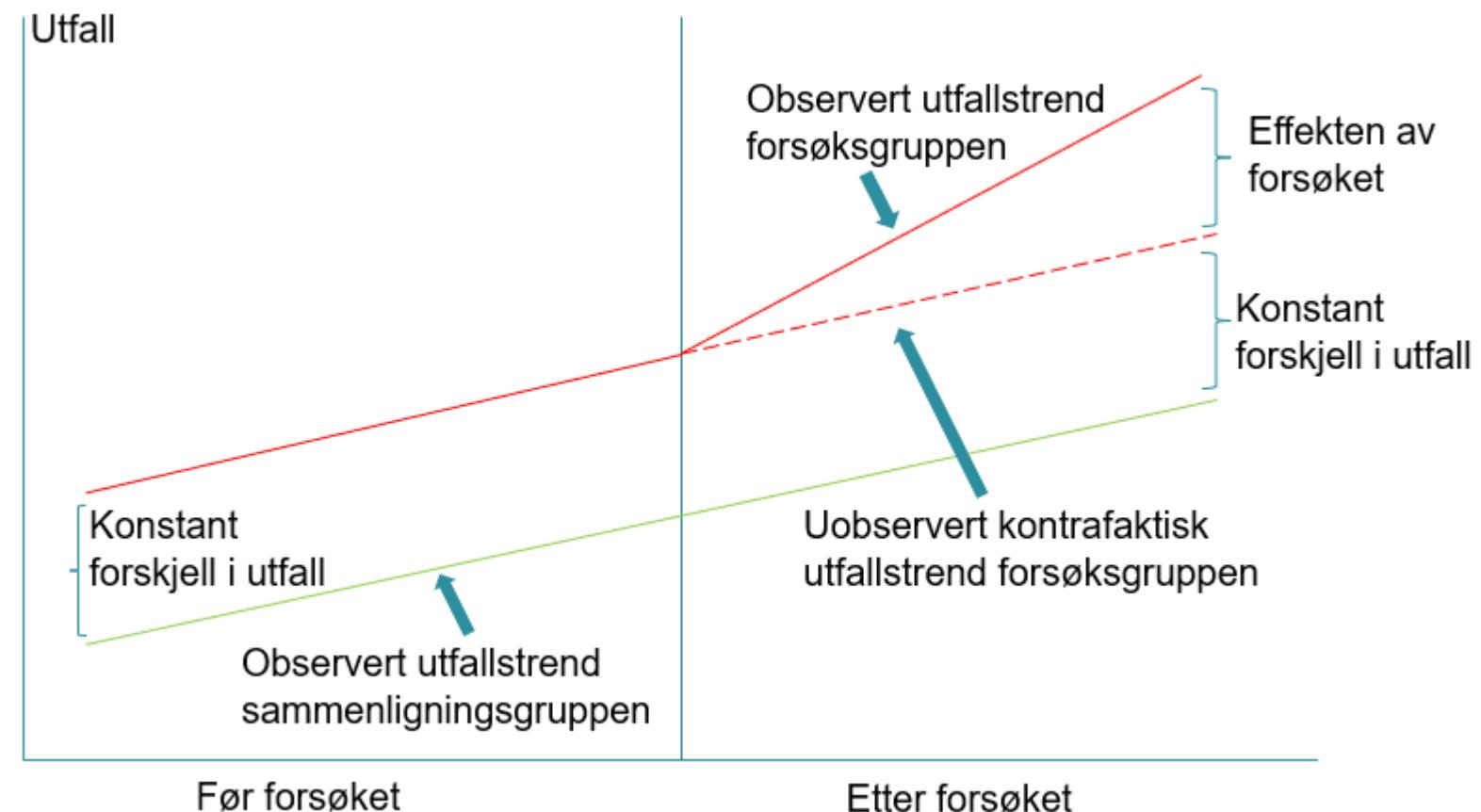
- Første differanse (treat vs. non-treat)

Sammenligne hver
gruppe før og etter en
endring

- Andre differanse (pre vs. post)

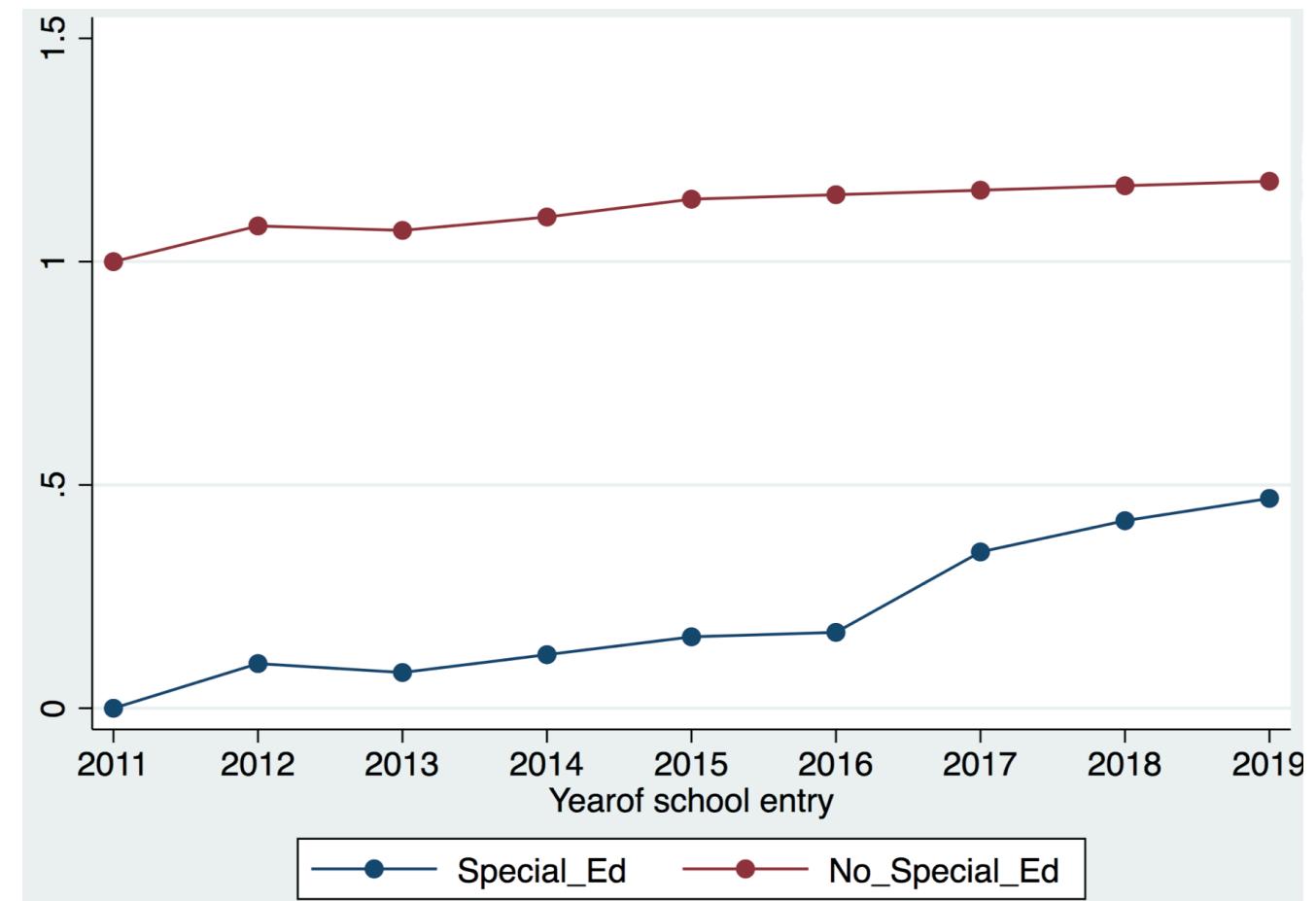
Antakelser:

- Ingenting annet skjer samtidig
- Felles trend



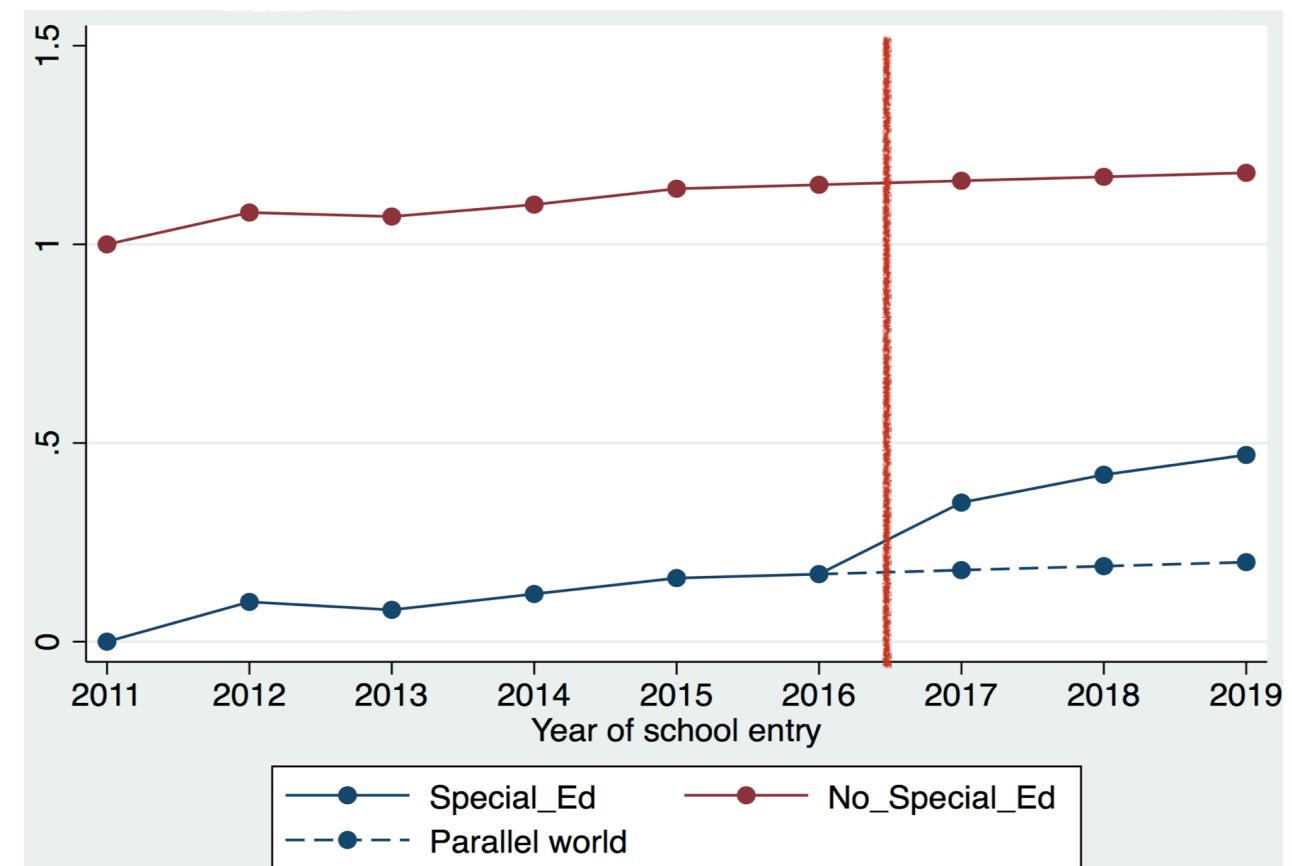
Kvasi-eksperimenter: Forskjell-i-forskjeller

- Forskjellen i resultat på kartleggingsprøve før og etter tiltak
- Forskjellen i resultat på kartleggingsprøve i for de med og uten vedtak om spesialundervisning



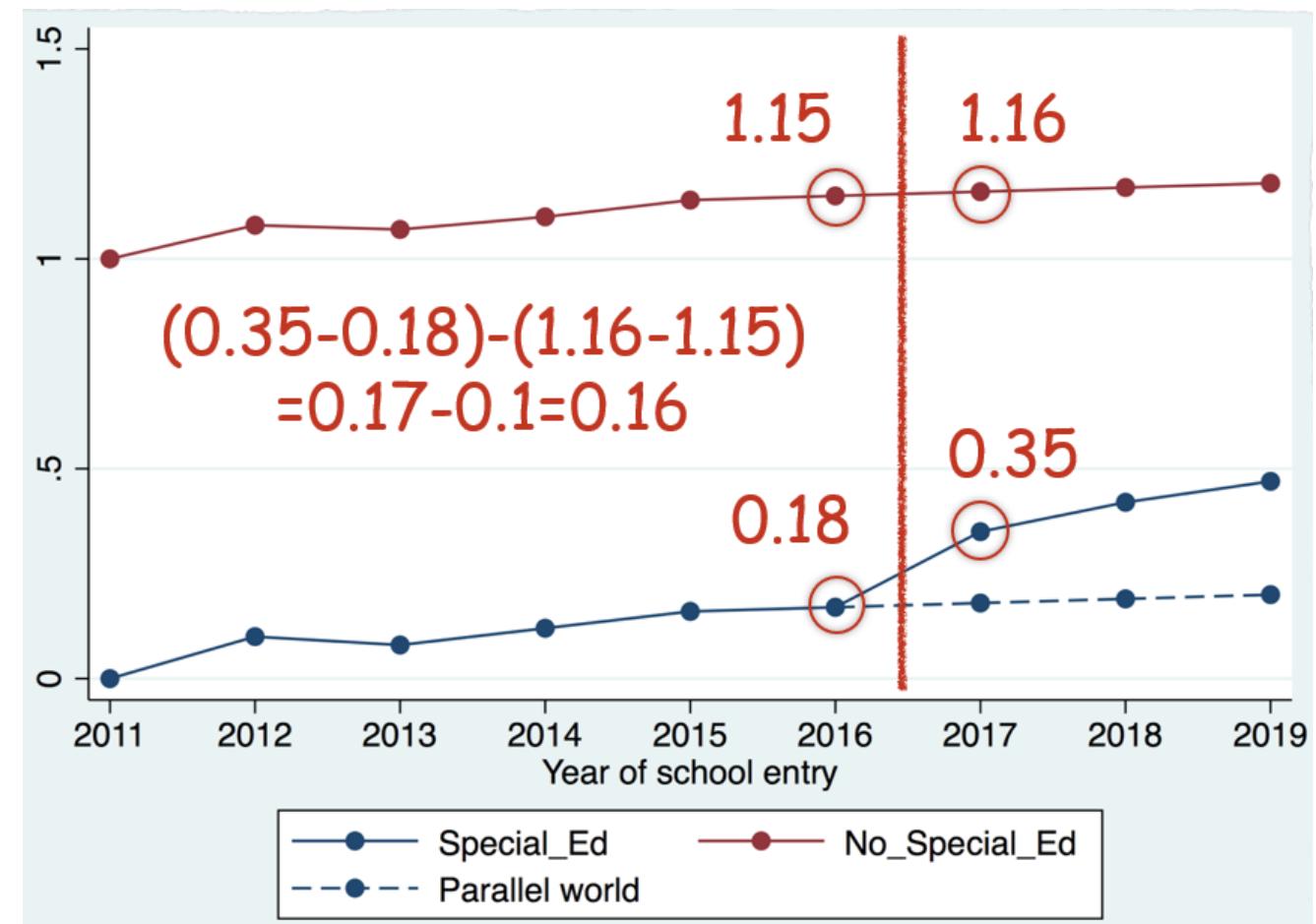
Kvasi-eksperimenter: Forskjell-i-forskjeller

- Forskjellen i resultat på kartleggingsprøve før og etter tiltak
- Forskjellen i resultat på kartleggingsprøve i for de med og uten vedtak om spesialundervisning



Kvasi-eksperimenter: Forskjell-i-forskjeller

- Forskjellen i resultat på kartleggingsprøve før og etter tiltak
- Forskjellen i resultat på kartleggingsprøve i for de med og uten vedtak om spesialundervisning



Utfordringer

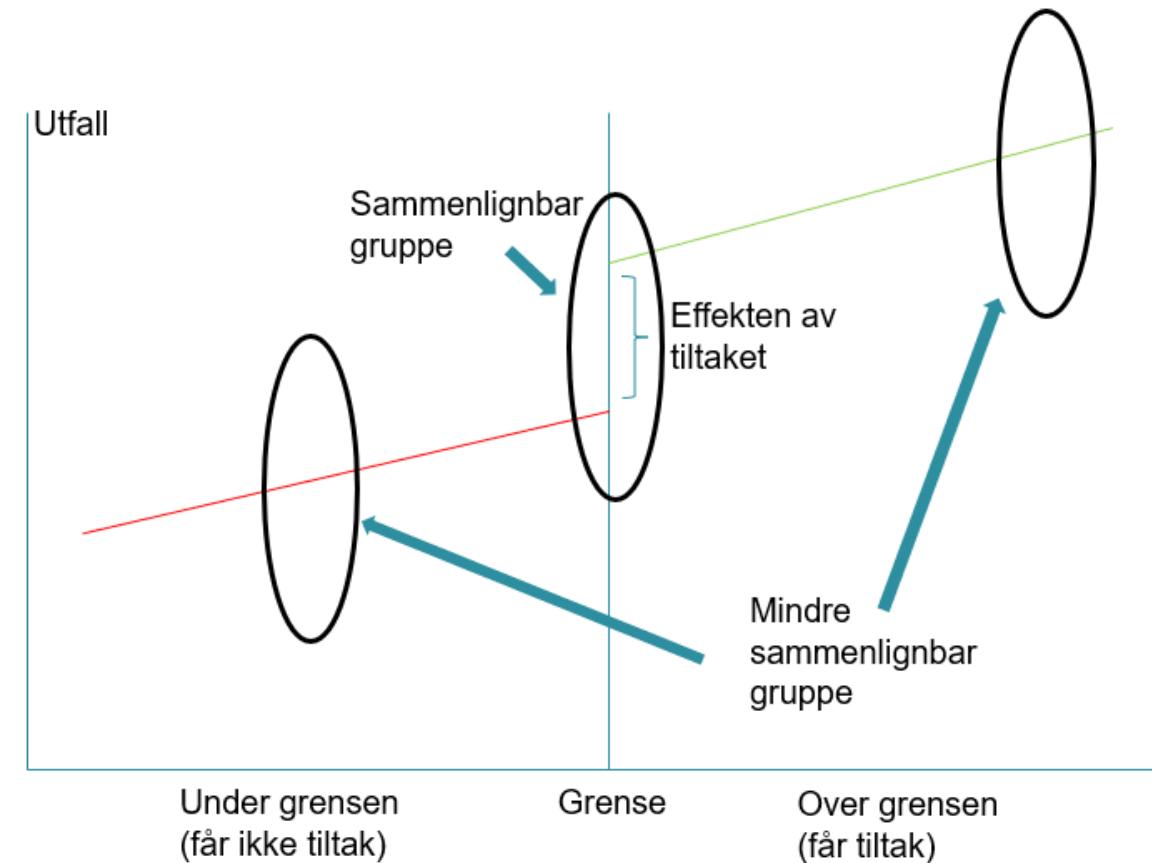
- Må kunne rettferdiggjøre at barn med vedtak før 1. august 2016 er sammenlignbare med barn som får vedtak etter 1. august 2016 (avbrutt tidsserie)
- Må kunne rettferdiggjøre at forskjellen mellom barn med og uten vedtak er sammenlignbare over tid før tiltak (felles trend) og at endringen etter tiltak skyldes tiltaket og ikke f.eks. endringer i gruppessammensetning (forskjell-i-forskjeller)
- Hvordan?
 - Lokal kunnskap
 - Logikk
 - Smart resonnering
 - Teste antakelsene sine

Kvasi-eksperimenter: Instrumentvariabel

- Finne en variabel (IV) som er korrelert med forklaringsvariabel (X) og ikke korrelert med andre variabler som påvirker det som skal forklares (Y)
- Instrumentet skal ikke ha noen effekt på Y utenom effekten den har gjennom X
- IV “renser” estimatet for andre påvirkninger og gir effekt av X på Y
- Mer sannsynlig å få tiltak om spesialundervisning hvis kommunen har høy PPT-tetthet
- Hvis PPT-tetthet ikke er relatert til noe annet som påvirker resultater på kartleggingsprøver, annet enn å øke andel som får tiltak om spesialundervisning, så kan dette være en IV-variabel
- Sannsynlig?

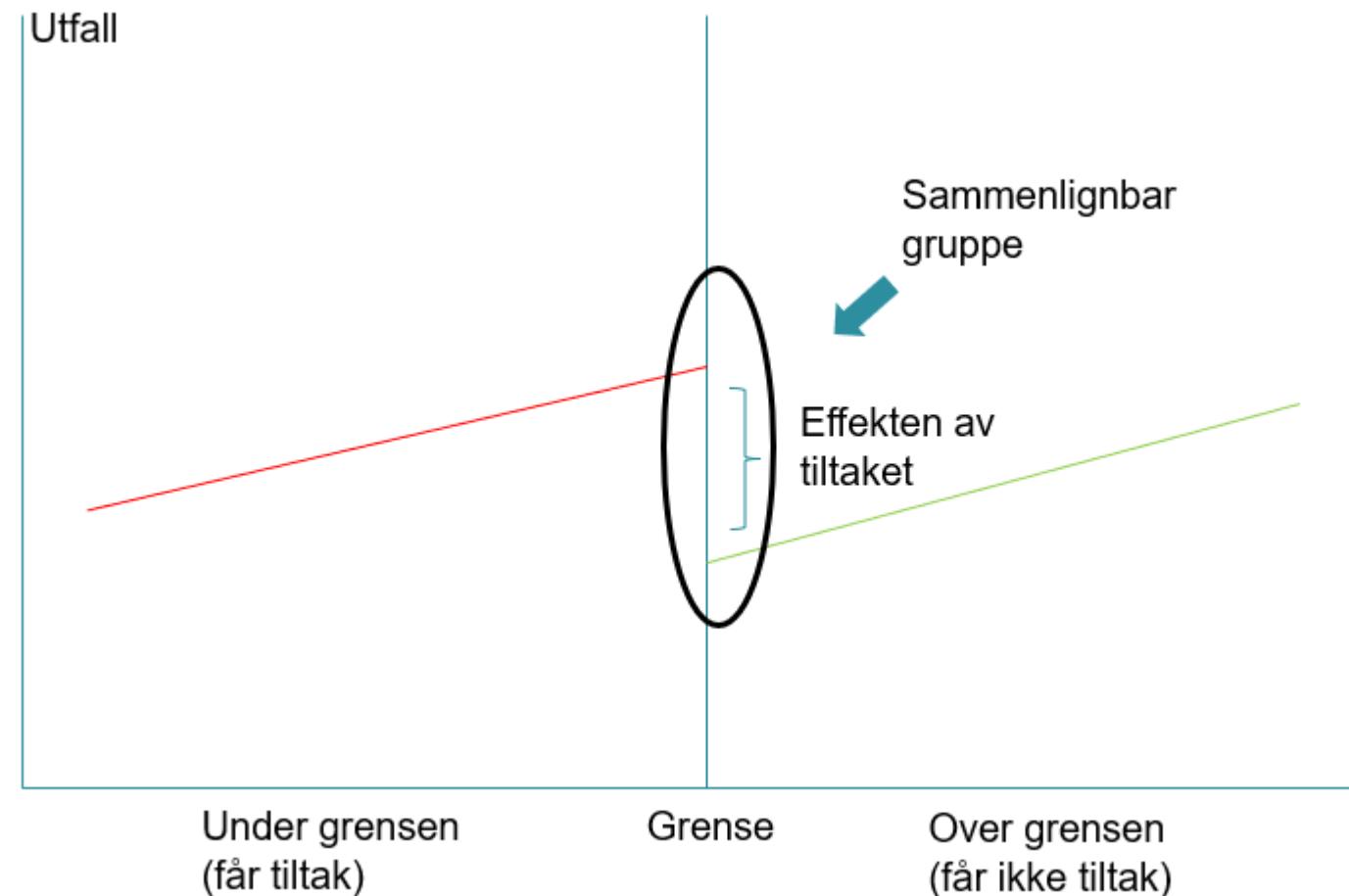
Kvasi-eksperimenter: Regresjonsdiskontinuitet

- Tiltak bestemmes av en grense
- Sammenligning av de som er rett over og rett under grensen gir effekten av tiltaket



Kvasi-eksperimenter: Regresjonsdiskontinuitet

- De som skårer under grensen på en pretest får tiltak om spesialundervisning
- Sammenligne de som akkurat kom over grensen med de som akkurat kom under



Kvasi-eksperimenter: Et eksempel



Child Development, xxxx 2013, Volume 00, Number 0, Pages 1–19

Little Evidence That Time in Child Care Causes Externalizing Problems During Early Childhood in Norway

Henrik D. Zachrisson
Norwegian Institute of Public Health
and The Norwegian Center for Child
Behavioral Development

Eric Dearing
Boston College

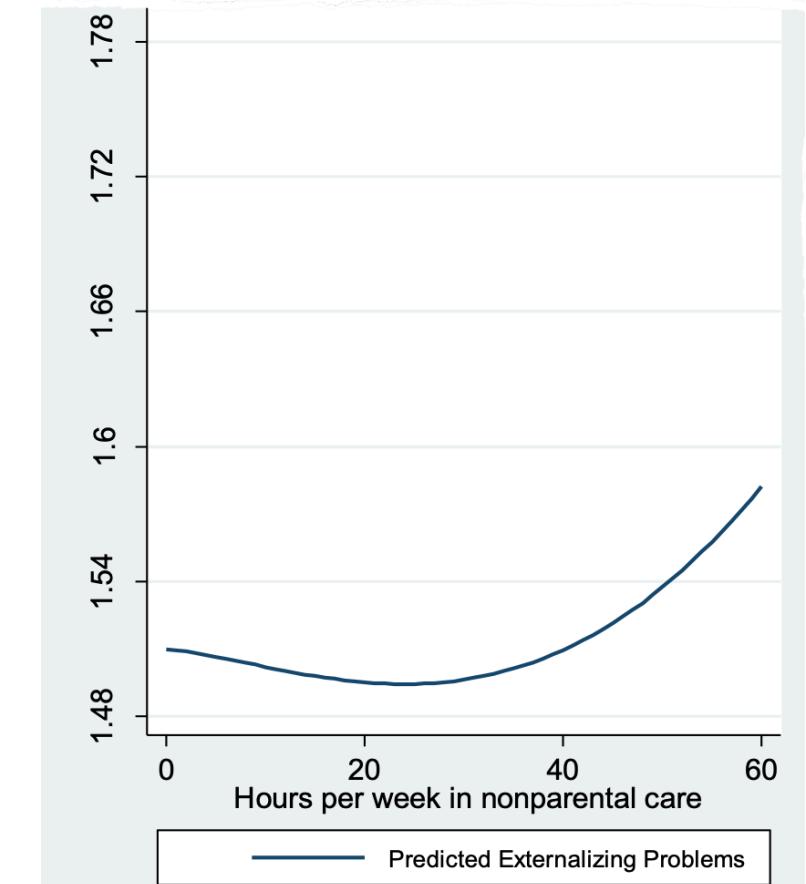
Ratib Lekhal
Norwegian Institute of Public Health

Claudio O. Toppelberg
Harvard Medical School and
Judge Baker Children's Center

- Dårlig samvittighet gir der jeg ikke utsette meg selv for. Ja, ja, da ble han sist i dag

Er det status å hente tidlig i barnehagen?

J annonse
I Hvafor vises denne annoncen?



Tredjevariabler/reversert kausalitet? Sammenligne søsken



Tredjevariabler/reversert kausalitet? Sammenligne med seg selv

Endring - uttagering



Endring - tid i barnehage

Take-home

- Kvasi-eksperimenter har ikke randomisering
- Må identifisere trusler- tenke smart, kjenne kontekst, kjenne data
- «When it comes to causal inference from quasi-experiments, design rules, not statistics» (Shadish & Cook, 1999) »

