

Hvordan kan vi få bedre kunnskap om hva som virker?

Seminar Finansdepartementet
Oddbjørn Raaum i samarbeid med Simen
Markusssen og andre kolleger



Stiftelsen Frischsenteret for samfunnsøkonomisk forskning
Ragnar Frisch Centre for Economic Research
www.frisch.uio.no

Umulig å vite...sikkert

- Vi kan aldri vite med sikkerhet at et "tiltak" (ikke) virker (: har intendert effekt)
- Men vi kan ha større eller mindre tiltro i en retning
 - Usikkerhetsmål
- Evalueringer bygger på antakelser ("identifying assumptions")
- Implementeringsdesign gjør at evalueringer kan bygge på svakere forutsetninger => mer pålitelige resultater

=> Flere forsøk/piloter før full gjennomføring

Tiltakseffekt kan være så mangt....

I. Effekt av hva?

(alltid sammenlikning av alternativer)

II. Effekt for hvem?

(ikke alt virker likt på alle)

(flere nivåer; "elever og lærere")

III. Effekt på hva?

(mange ulike utfall)

Effekt og kontrafaktiske utfall

Effekt (av A vs B) = utfall ved A - utfall ved B

Observerer aldri begge utfall for samme individ => effekt er et teoretisk begrep. Kan aldri observeres!

Standard opplegg i effektevalueringer:

Individer som faktisk opplever B brukes til å anslå hva de med A ville ha oppnådd under dette alternativet (kontrollgruppe-ideen)

Ideal: Naturvitenskapelige, medisinske eksperimenter

Antakelse: Et individs plassering i A eller B påvirker ikke andres personers utfall

(Gjennomsnitt mange individer for å unngå tilfeldigheter, hensyn til observerbare forskjeller)

Hvordan etablere data om kontrafaktiske utfall?

- Innen "individ " (før-etter)
 - Etter vs før. A: etter, B: før.
 - Utfall 'før' representerer kontrafaktisk utfall
 - Ulempe; andre forhold endrer seg over tid
- Mellom "individer" (kontrollgruppe)
 - Utfallet i gruppen som opplevde B brukes til å måle kontrafaktisk utfall for gruppen med A.
 - Ulempe; er det bare tiltaket som skiller de som opplever hhv A eller B?
- Kombinasjon av innen og mellom "individer" (følger begge grupper over tid)

Hva er problemet med ordinære data?

- Seleksjonsproblemet
 - De som 'velger'/blir valgt ut til A er oppnår andre utfall enn de med B, av andre grunner enn forskjeller i "miljø" som følge av A eller B.
- Hva kommer det av? Ofte(st) bevisste valg, avhengig av omstendigheter
 - Ansatte i offentlige sektor (lærere, helsepersonell etc)
 - Klienter (arbeidsledige, elever etc)

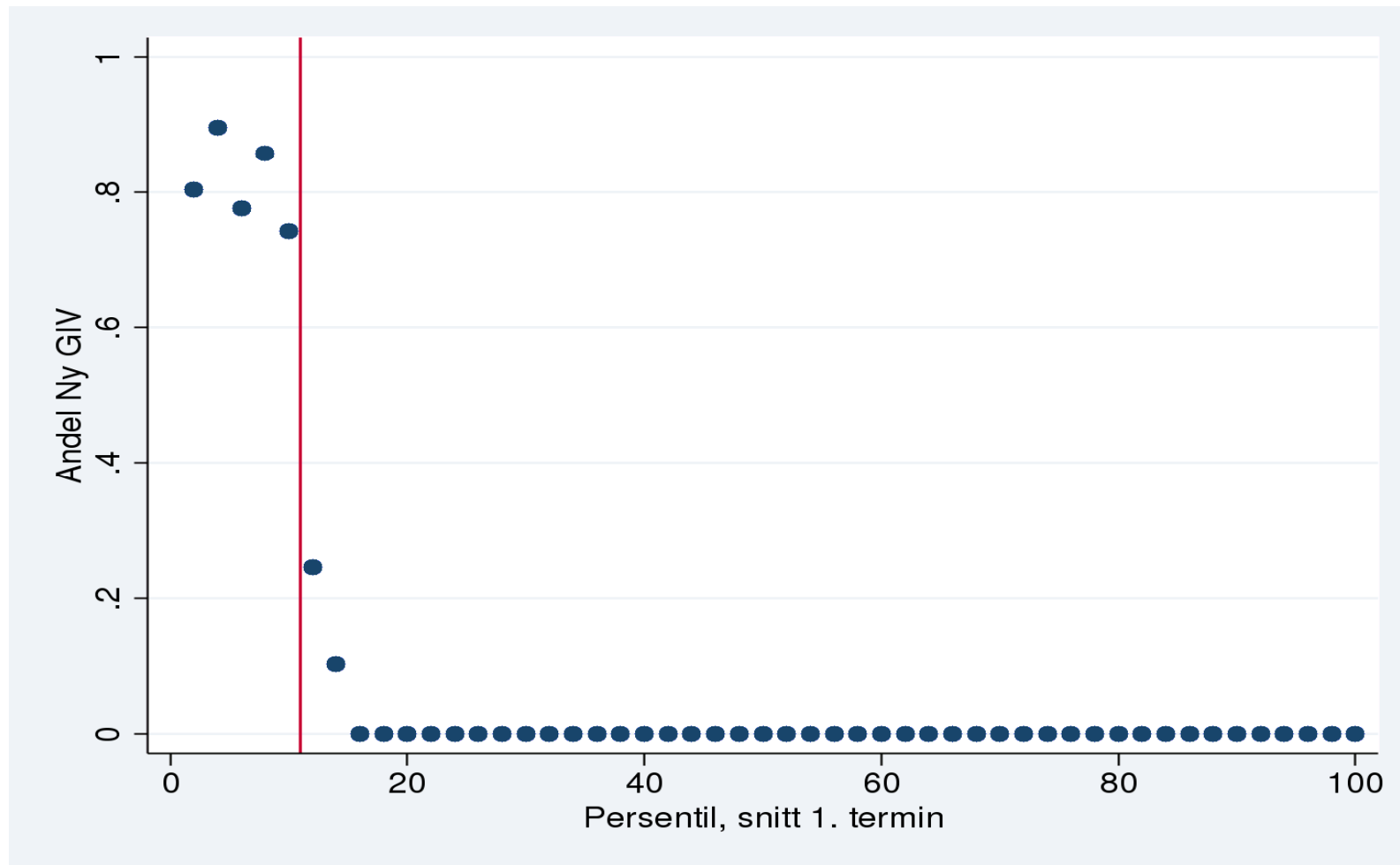
Løsningen:

Implementering med skarp tilordningsregel

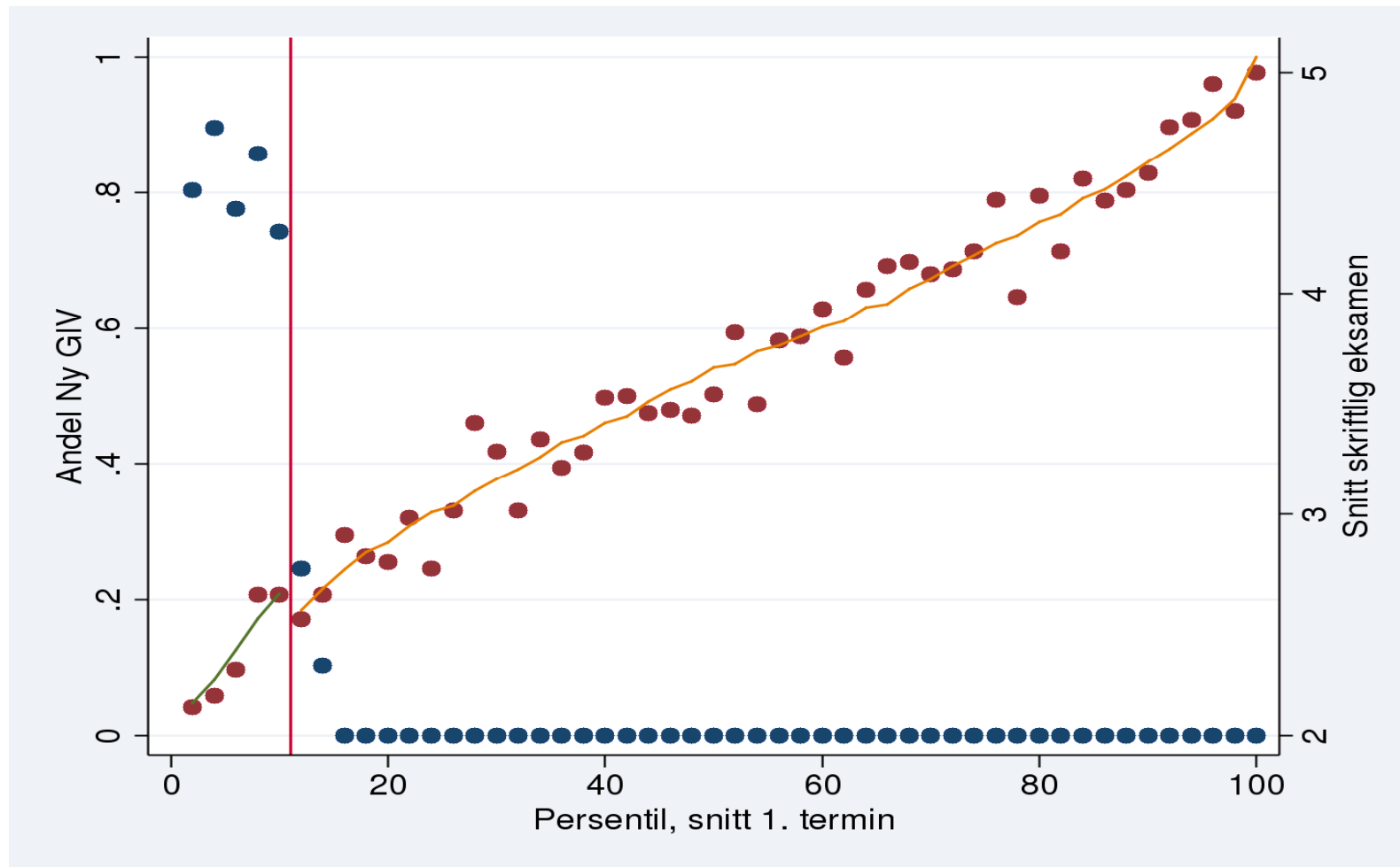
- Kjente sannsynligheter (loddtrekning, RCT)
 - Kontrollerer sannsynligheten for A og B
 - Ikke nødvendigvis lik sannsynlighet for alle

Eksempler: Arbeidsmarkedspolitikken; AMO kurs 1991, nye initiativ pågående, langt mer utbredt i andre land
- Objektiv, skarp, ikke-påvirkbar tilordningsregel (regression discontinuity, RD)
 - Vet (på tilfeldigheter nær) hvem som 'akkurat får' og hvem som 'akkurat ikke får'
 - Bygger på at individer rundt denne marginen er like
 - Eks: Permisjonsrettigheter

Ny GIV-intensivopplæring 10.klasse-elever (i Stavanger)



Ny GIV og skriftlig eksamen - Stavanger



Hvorfor er RCT og (intendert) RD så lite utbredt?

- I. Er "gullstandarden" oppskrytt ?
 - Alvorlige svakheter?

- II. Praktiske, moralske og juridiske hindre
 - Ulempene dominerer gevinsten? Kan leve med å "anta" at tiltak virker sånn nogenlunde?

I. Tolkning av effekter identifisert ved RCT og RD

a. Ikke-deltakere upåvirket?

- Relative forskjeller innen gruppe
 - Som ved loddtrekning (RCT)
 - Effekt på marginal elev – effekt av utvidelse
- Gjør Ny GIV-elevene på skolen det bedre, eller gjør de andre det dårligere?
 - Eksempel: Superlærer til Ny GIV? Konsekvenser for de andre?
- Problem: Bygger på antakelsen om at "treatment" kun har effekt på de som mottar den

b. Finner vi effekten av interesse?

RCT: Gjennomsnitteffekten ved loddtrekning. Med heterogene effekter; interessert i effekter for dem som blir påvirket av politikkendring (får A i stedet for B) ('tiltaket benyttes av dem som har størst utbytte av det')

c. Non-compliance (ikke alle tar i mot tilbudet / "tar pilla")
=> effekten av tilbud . En slik 'intention to treat'-effekt kan skaleres opp til gjennomsnittseffekt for deltakere hvis vi kjenner andel som "tar pilla"

II. Praktiske hindere

- Urettferdig? Ryggmargsreflekser i opinionen
 - Loddtrekning er urettferdig, selv ved fordeling av et knapt gode
 - Alle bør få, selv om vi er usikre på om det virker
 - Innen utdanning: "Eksperimenterer med barnas framtid"
- Juridiske rettigheter
 - Lovverk og forskrifter => individuelle rettigheter
 - Krever hjemmel for å endre tilordningspraksis
- Kompensasjon
 - Ødelegger designet (med mindre det vil bli en del av politikken)
 - Alt annet er ikke likt
- Sabotasje i "sektoren"
 - Fagfolka vet hva som er best
 - Skader egeninteresse?

Flere problemer....

- "Overløpere" (Cross-overs)
 - Studieobjekter bytter status fra A til B (kanskje uten at vi vet det som utenforstående, forskere)
- Forsøksstatus har en direkte effekt på adferd
 - Hawthorne-effekter
 - Herunder placeboeffekter
- Tilpasninger / endring i adferd på "andre områder" / kompensasjon
 - Endring i andre forhold som respons på A eller B
 - Avgjørende om dette vil skje ved evt senere implementering
- Offentlige innkjøpsregler
- Personvern

Tidligere/pågående ekse

- Loddtrekning til arbeidsmarkedskurs i Norge 1991 (Raaum og Torp, 2002)
- NAV : Effekter av ventetid til tiltak (pågående)
 - Fire fylker, et utvalg attføringstiltak (bl.a. arbeidspraksis i skjermet virksomhet)
 - Etter tiltak er bestemt trekkes et lodd. Loddet avgjør om en havner på topp eller bunn av ventelisten (20% gis prioritet, dvs. settes på toppen)
 - Dette gir en tilfeldig tildelt forskjell i ventetid før en kan begynne på tiltaket
 - Under gjennomføring nå

- Vi vil så evaluere effekter, f.eks hvorvidt det påvirker sysselsetting m.m. f.eks 24 mnd etter en ble satt på venteliste
- Forsøket vil kunne gi svar på hva som ville vært effektene av økt tiltakskapasitet (dvs. redusert ventetid). Dette kan undersøkes for hvert enkelt av tiltakene som inngår.

Tidligere/pågående II

- NAV : Forsøk med nytt tiltak: Supported Employment (SE) – ikke startet ennå
 - NAV/AD vil undersøke hvorvidt SE er bedre enn dagens to oppfølgingstiltak (Arbeid med bistand AB og Oppfølgingstiltaket OT).
 - Etter at det er bestemt at brukeren skal ha oppfølging (AB/OT) trekkes lodd om hvorvidt brukeren i stedet skal ha SE. Dersom det ikke blir SE avgjør saksbehandler (som i dag) hvorvidt AB eller OT skal benyttes
 - SE sammenlignes med «business as usual». Forsøket gir svar på hvorvidt dagens regime (AB og OT) skal erstattes av SE
 - Mange komplikasjoner, bl.a. knyttet til innkjøp av tiltakene



Ideer til nye tiltak

- Lærerløftet
 - 10 000 (?) lærere etterutdanning over 5 år
 - Ikke alle kan få på en gang; ulike faser
 - Team/enkellærer velges ut ved RCT eller RD til ulike faser
 - Elevutfallstudie krever kobling lærere og elever; sammenlikne kull på tvers av skoler
 - Nivå: Kommune for grovt
 - Må bryte dagens praksis
 - Organisering: Fri fra timer eller på fritida? Trekke lodd mellom alternative tilbud om organisering

- NAV

- Det er svært store variasjoner i hvor mange brukere en NAV-veileder skal følge opp (KVP: 18, AAP: 83 (snitt), varierer mellom 30 og 300 mellom kontor).
- Mulig forsøk: la loddrekning avgjøre først hvor mange klienter hver veileder skal ha, deretter hvilke brukere som skal til hvilke veiledere
- Eksempel: på et kontor med tre veiledere kan disse da ha hhv 50, 100 og 150 brukere hver. Vi kan så studere utfall x mnd etter eksperimentet, f.eks sysselsetting, overgang til uføretrygd, fortsatt AAP-mottak osv.
- Vil svare på hva som ville vært effekten av å øke bemanningen i NAV for oppfølging av AAP-mottakere.

Randomiseringsnivå

- Person
- Organisasjon
- Region/kommune

Implikasjoner for praksis i Norge

- Ved innføring av nye tiltak bør en tidlig tenke på hvordan en skal kunne besvare spørsmålet om det virker! Må planlegge erfaringsmaterialet man skal bruke
 - Vente med fullskalaimplementering? Eksempel lærerløftet
- Etisk forsvarlig når vi ikke vet hva som er best? De skråsikre trenger ikke eksperimenter
- Eksperimenter krever ulikebehandling
 - Lettest ved innføring av nye tiltak, piloter?
- Velge A og B som i forventningsforstand er like gode?
 - Hverken A eller B trenger å være status quo

- Forutsetter politisk klarering og rakryggede politikere som står fast når medieulvene uler
- Implementeringsdesign:
 - Avveining mellom praktisk-juridiske problemer og gevinsten ved å kunne gi svar på hva som virker
 - Designe slik at effekter avspeiler politikkendringer man ser for seg. Effekter for dem som påvirkes av politikkendringer => kreativ design, må kjenne "complier" gruppen
- Invitere forskere når reformer skal iverksettes
 - Ydmykhet i forhold til hva vi vet (eks lærerkompetanse og elevutfall) og hva som kan implementeres