



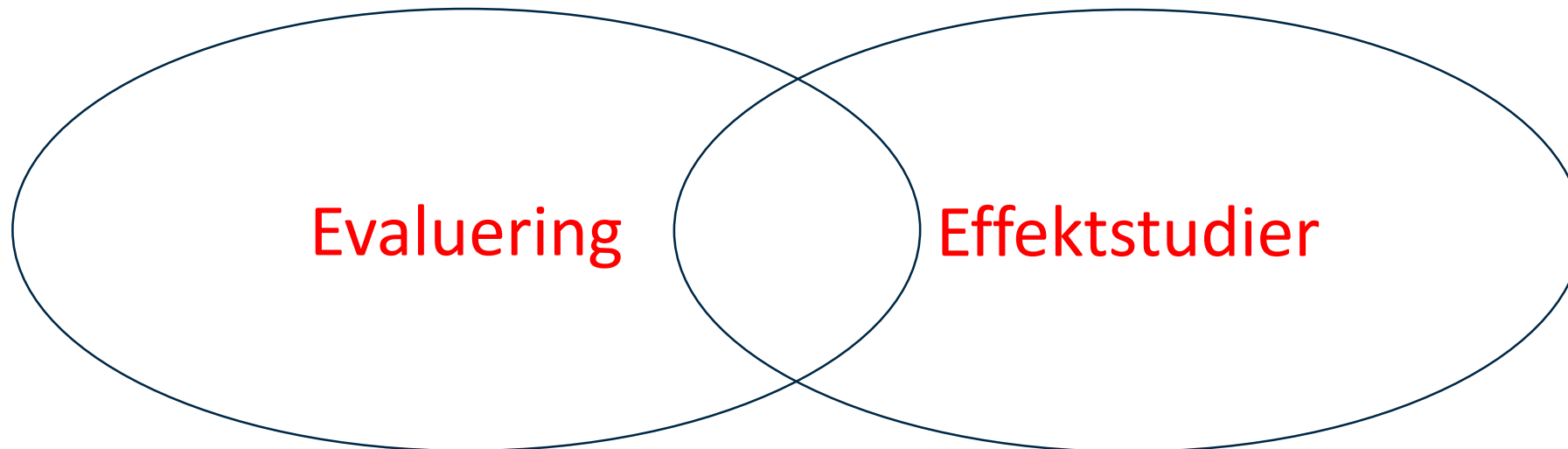
# - JA, MEN HVA ER EFFEKTEN? OM DESIGN OG FORTOLKNING AV EFFEKTSTUDIER

Håkon Finne, SINTEF

Evalueringskonferansen 2018-09-21, Kristiansand

# Effektstudier og evaluering

---



- Kunne ha listet opp ingredienser, men går litt dypere inn i det
- Ikke spørsmål om en bestemt teknikk eller metode!

# Disposisjon

---

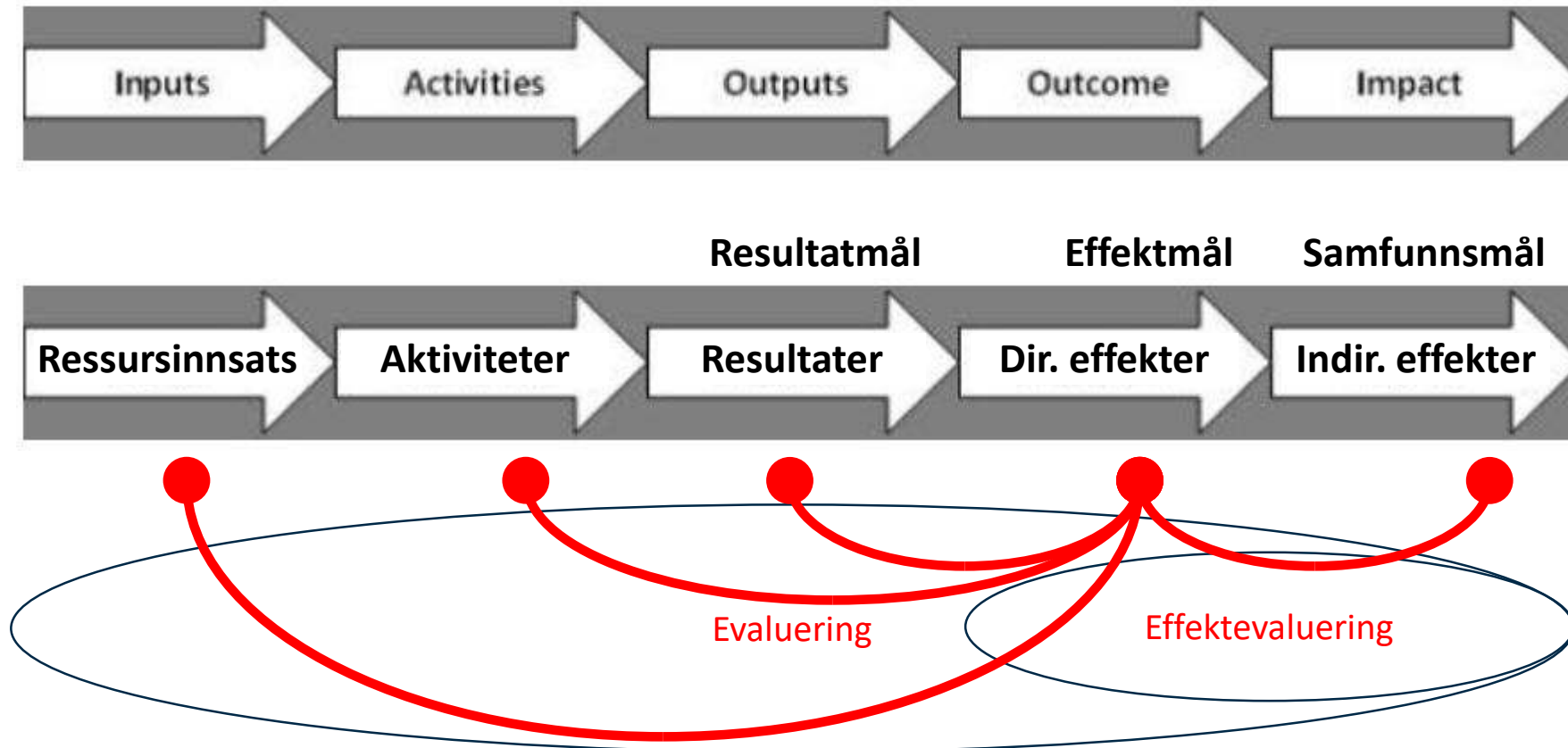
- Problemforståelse
- Programlogikk som basis for evaluering og effektstudier
- Tre hovedstrategier for effektstudier, etter fenomenets kompleksitet
- utfordringer ved noen sviktende forutsetninger, med spesiell fokus på randomiserte kontrollerte studier (RCT)
- Konklusjon

# Hvori består forvirringen, hvorfor trengs denne forelesningen?

---

- Stigende etterspørsel etter evidens for effekter av tiltak, etablert praksis for å bestille eksterne evalueringer av tiltak
- Feilaktig oppfatning av at effektmåling er en bestemt type statistisk metode
- Er effektevaluering det samme som effektstudier?
- Ønske om entydige svar, men ofte vanskelig å tolke resultater av effektstudier entydig
- Også: Økende interesse for å forstå underliggende sammenhenger ("hvordan virker det?")

# Programlogikk for evaluering



# Tre hovedstrategier for å undersøke effekter

---

- Må skjønne hva som skjer
- Må kunne attribuere endringer i utfall til endringer i tiltak – forholdet til kontekst er avgjørende
- Må forstå – og kommunisere – forutsetninger og antakelser man gjør, og konsekvenser av dem
- Valg av generell tilnærming (strategi, design, metode, teori, grundighet, ...) avhengig av mange ting, ikke minst ressurser og en form for kost/nyttevurdering
- Valg: innenfor eller utenfor en evalueringsramme

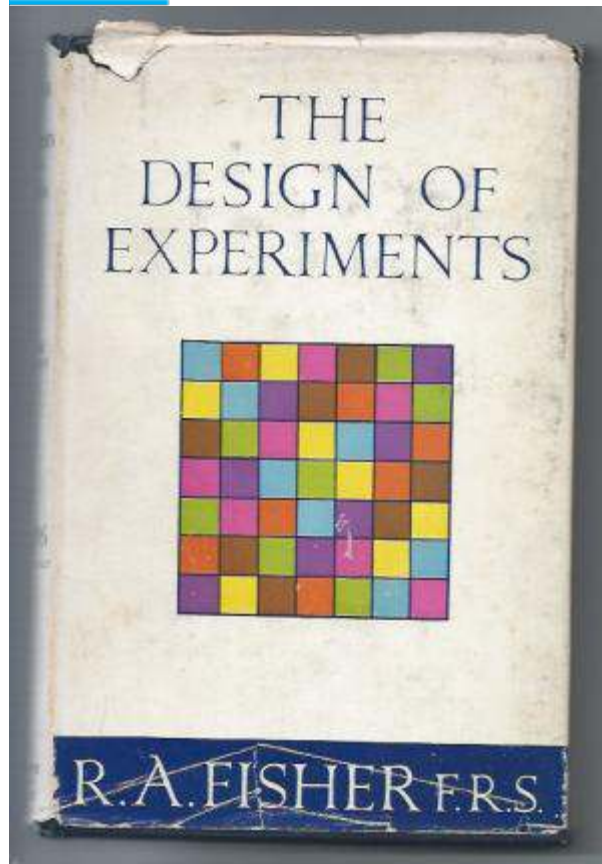
# Strategi 1: Isolere fenomenet, låse konteksten

---

- Laboratorieeksperimentet:
  - Låser fast konteksten
  - Variierer input kontrollert
  - Måler output
  - Replikerer forsøket, korrigerer statistisk for målefeil
- Forutsetter at den kausale sammenhengen mellom input og output er uavhengig av andre påvirkninger av output, og bare ett ledd i den kausale kjeden
- Sjelden mulig strategi i sosiale system



# Strategi 2: Randomisere konteksten



Fisher (1935, 1971)

- Randomisert kontrollert forsøk (RCT): Strategi for å korrigere for kontekst som ikke lar seg låse fast
- Varierer input kontrollert  
Måler output, også i "kontrollgruppe"
- Medisinsk variant: "Double blind"
- Forbehold om at den kausale sammenhengen mellom input og output er uavhengig av andre påvirkninger av output, og bare ett ledd i den kausale kjeden
- Har litt ufortjent oppnådd status som gullstandard også for sosiale system



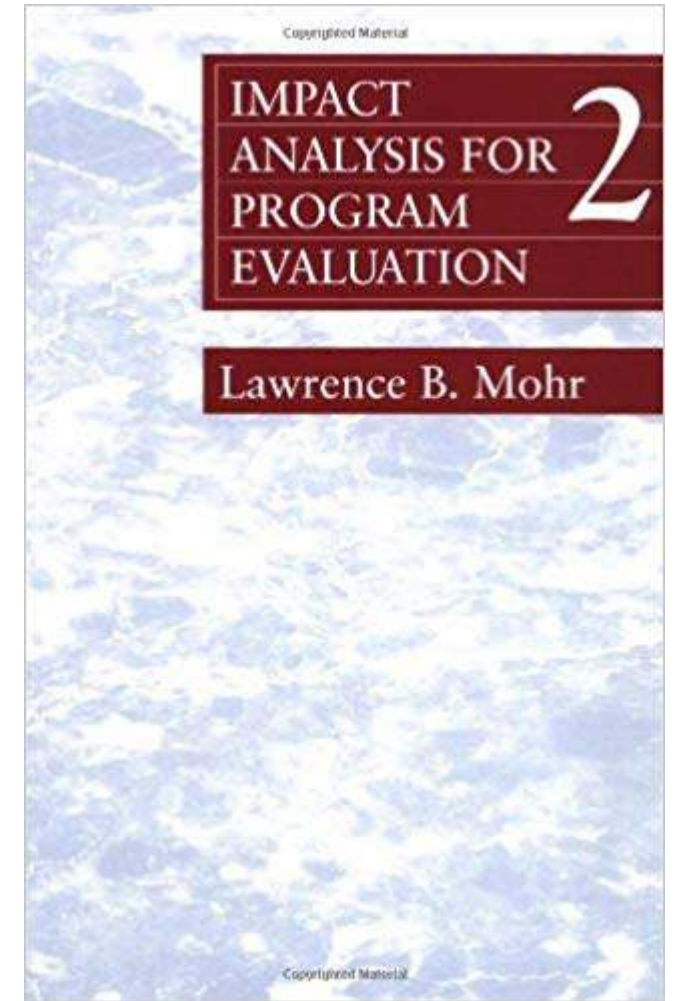
Bladbete (Wikipedia)



**Table 4.1** Characteristics of Design (placed here for ease of reference)

**Table 4.1** (Continued)

Design				Threats to Internal Validity						
Number	Chapter	Name	Diagram	Locus; Mode of Selection	Spuriousness/ History Time Order	Selection P	Q	Contamination		
1	4	one-shot case study	$T Y$	none						
2	10	ex post facto	assorted	decentralized	x	—	x	x		
Elementary quasi-experimental designs										
3	4	before-after	$A/C: Y_1 T Y_2$	centralized; autonomous/controlled	—	x	—	x		
4	4	comparative posttest	$(A): T Y_E$ $A: Y_C$	centralized; autonomous	—	—	x	x		
Quasi-experimental designs										
5	7	comparative change	$(A)/C: X_{1E} T Y_E$ $A/C: X_{1C} Y_C$	centralized; autonomous/controlled	—	—	—	x		
6	9	interrupted time series	$A/C: Y_1 \dots T Y_{T+1}$	centralized; autonomous/controlled	—	x	—	x		
7	9	comparative time series	$(A)/C: Y_{1E} T Y_{(T+1)E}$ $A/C: Y_{1C} Y_{(T+1)C}$	centralized; autonomous/controlled	—	—	—	x		
Fortified designs										
8	8	criterion population	$(A)/C: X_{1E} T Y_E$ $A/C: X_{1CP} Y_{CP}$	centralized; autonomous/controlled	—	—	—	x		
9	11	subjective	assorted, based on $T S Y$	centralized; randomized or autonomous/controlled	—	(at best)	(at best)	x		
Experimental designs										
10	6	regression discontinuity	$A/C: A_E T Y_E$ $A/C: A_C Y_C$	centralized; quasi-randomized	—	—	—	x		
11	4	R-comparative posttest	$R: T Y_E$ $R: Y_C$	centralized; randomized	—	—	—	x		
12	4	R-comparative change	$R: X_{1E} T Y_E$ $R: X_{1C} Y_C$	centralized; randomized	—	—	—	x		



# Strategi 3: Forstå fenomen i kontekst

**A certain mother habitually rewards her small son with ice cream after he eats his spinach. What additional information would you need to be able to predict whether the child will: a. Come to love or hate spinach; b. Love or hate ice cream, or c. Love or hate Mother?** Bateson (1972)



Spinat (Wikipedia)

- Primærinteresse: Den kausale sammenheng mellom iskrem og spinat, men sterk vekt på å forstå både andre samspill, "bivirkninger" og prosesser
- Bredt spekter av mulige forskningsdesign, kvalitative og kvantitative
- Ingen forutsetninger om at den kausale sammenhengen mellom input og output er uavhengig av andre påvirkninger av output, eller retning

# Virkningsanalyser (strategi 3)

---

- Komplekse sosiale systemer er normalt tilstanden, men noenlunde stabile mønster forekommer ganske ofte, og når disse er forstått, kan strategi 2 være et realistisk alternativ
- Hvordan virker hva for hvem, under hvilke betingelser
- Hvilke dynamikker spiller inn, wicked problems, positive og negative tilbakekoblingsløyper, årsak og virkning kan bytte plass, ...
- Betrakte trekantforholdet kontekst, mekanisme og direkte effekt (outcome) i stedet for forholdet mellom input og direkte effekt i en kontekst (Pawson, real/ist/ic evaluation)

# Bør evalueringer følge hele kjeden?

---

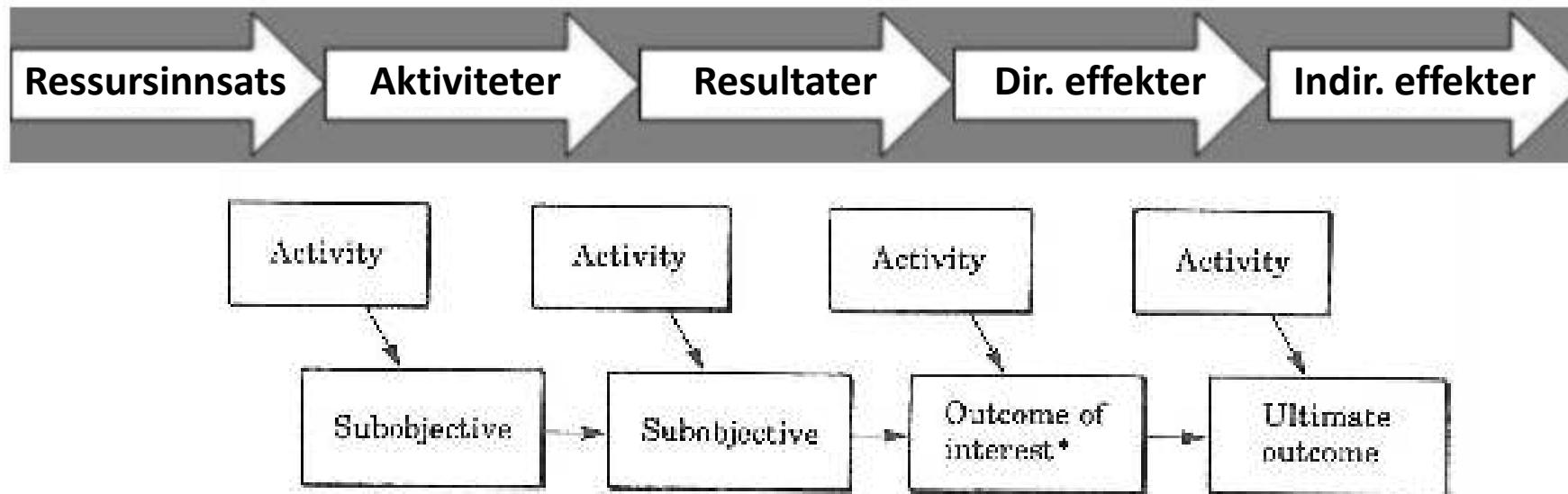


Figure 2.1. Outcome Line

\*Universal; must appear in all evaluations.

Mohr (1995)

# Hvor langt bør en evaluering strekke seg?

---

- Man må ha tiltro til at resonnetet for resten av kausalkjeden holder noenlunde mål
- Fare 1: Et MRS-opplegg som fokuserer for mye på lett oppnåbare resultater som ikke nødvendigvis bidrar til mål lengre ut i kjeden
- Fare 2: At kausalkjeden ikke holder mål

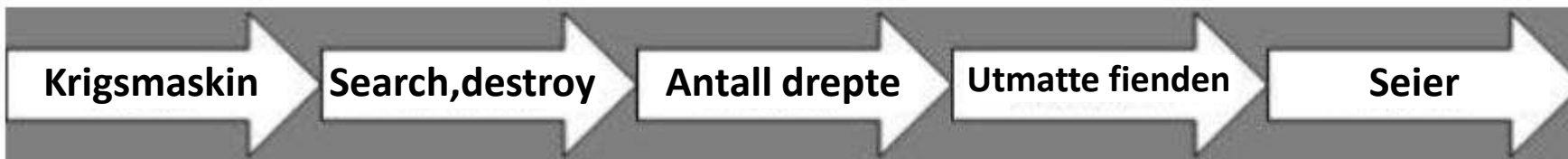


# Feilslått MRS: Westmorelands Vietnam-krig

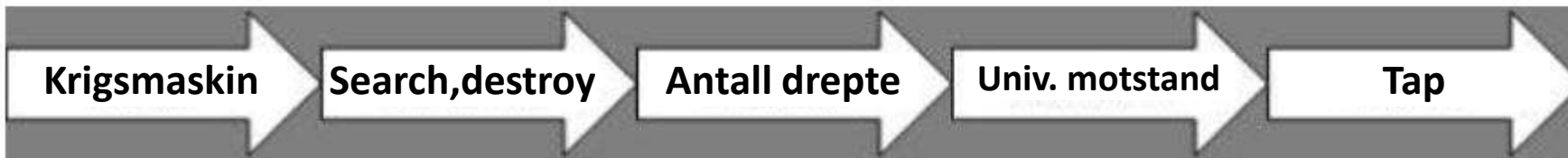
Klassisk territorialkrig:



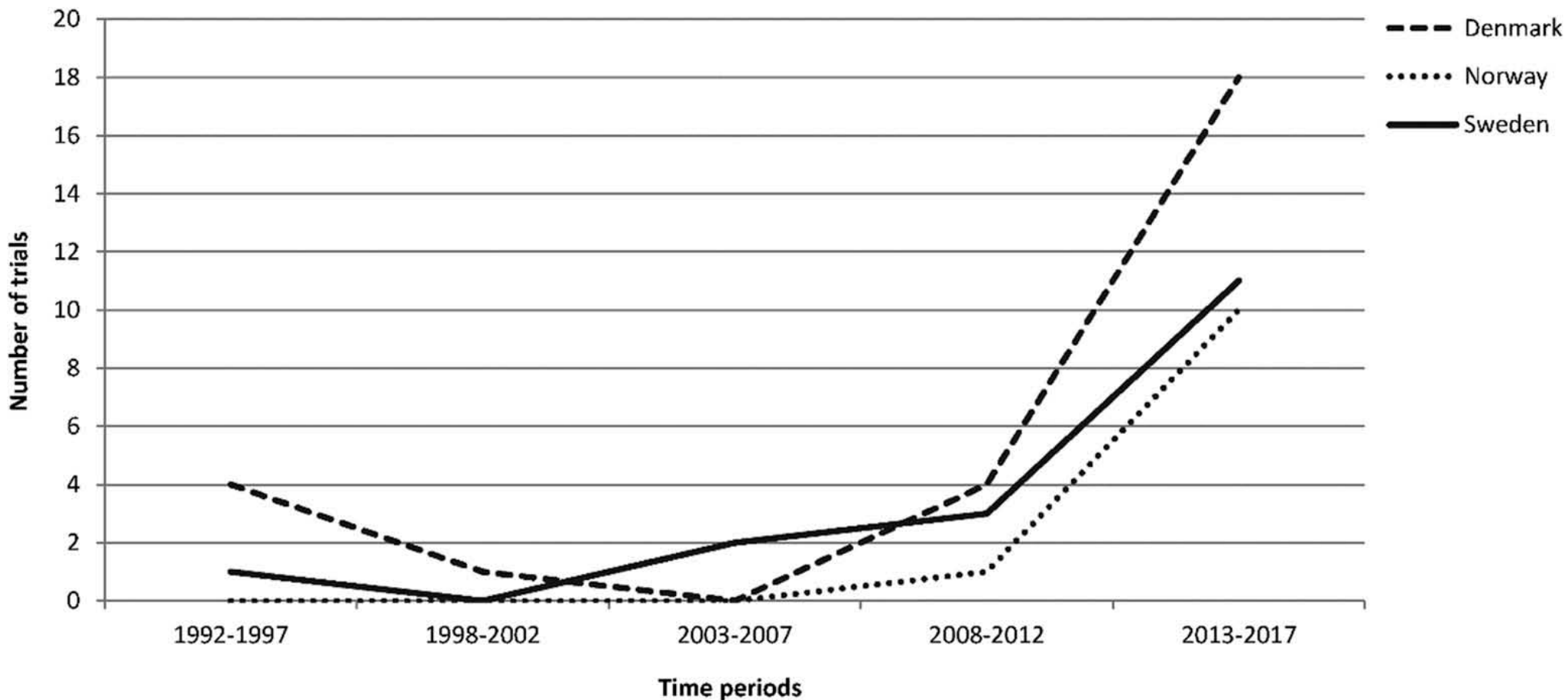
Krig mot gerilja infiltrert i terreng og samfunn:



Langvarig fokus på resultatindikator, bieffekt viktigere enn tilsiktet effekt:



# Randomiserte kontrollerte studier i utdanningsforskning



Kilde: Pontoppidan et al. (2018)

# Norgeshistoriens største RCT

## Meld. St. 21

(2016–2017)

Melding til Stortinget

### Lærelyst

– tidlig innsats og kvalitet i skolen

Stortinget har vedtatt å bruke nærmere 500 millioner kroner over fire år for å finne effekten av nye måter å bruke ekstra lærerstillinger i skolen. Opp mot 400 millioner skal gå til ekstra lærerstillinger på 1. – 4. trinn i rundt 300 norske skoler. Femti millioner skal gå til forskning på hvilke effekter forsøkene får.





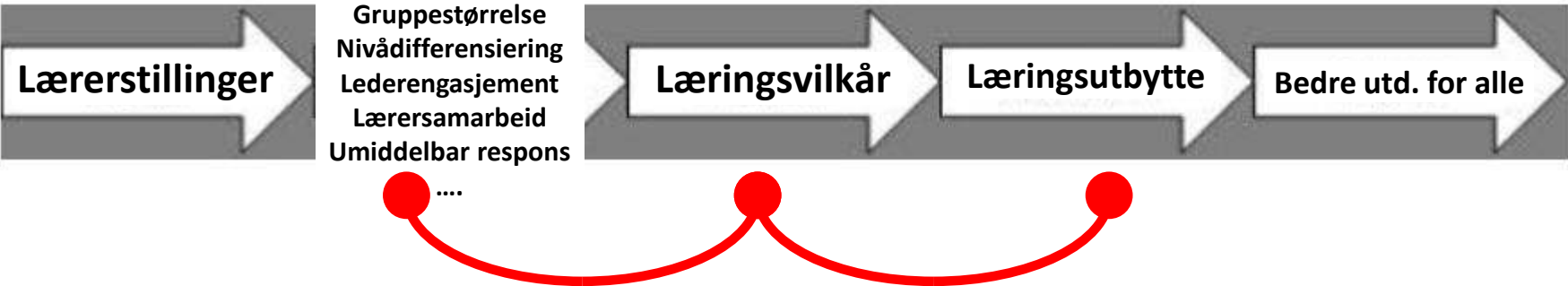
# Forskernes hypotese:

## At effekten avhenger av hva skolene og lærerne gjør med ekstraressursen

Bevilgende myndigheter, forenklet interesse:



Forskere:



# Utfordring: randomisering av input kan nulles ut

- Effektstudie forutsatte fire års virketid
- RCT-design forutsatte randomisert tildeling av lærerressurs i forsøkets virketid
- Politisk vedtak om lærernormen gir alle samme lærerressurs etter to år
- Forutsetningen for RCT-designets styrke er borte

## Lærernormen stopper skoleforskning

**POLITIKK.** Et stort forskningsprosjekt med 6.000 elever fordelt på 150 skoler blir innstilt på grunn av lærernormen.



Elevene på Borgen barneskole i Sarpsborg får fortsette å ha matteundervisning i små grupper, mens 6.000 elever på et annet prosjekt tas ut av forskningen til sommeren. Foto: Liv Dolmen

Berit Almendingen

18.4 2018 10:56



– Det var jo uventet at endringen kom underveis i prosjektet. Det var planlagt at tiltakene skulle ha en varighet på fire år. Nå blir det redusert til to, sier førsteamanuensis Oddny Judith Solheim ved Universitetet i Stavanger.

# Utfordringer ved effektstudier av programmer for bedriftsutvikling

---

- FRAM-programmet (fra ca. 1990): Lære fokusert målstyring mot forbedret bunnlinje
- Utfordring 1: Mange strategier, mange preferanser, mange motiver for deltakelse (Rolfesen mfl. 1994, Clausen 2003)
- Utfordring 2: Forskjellige strategier (vekst, nye produkter, kostnadsreduksjon osv.) har forskjellig forventet tidsforløp for bunnlinjeutvikling
- Utfordring 3: Mange andre forhold i og rundt bedriften som påvirker bunnlinjen

# Utfordringer ved effektstudier av bedriftsorientert prosjektstøtte til FoUol

---

- Tre komplementerende forklaringsmekanismer, tre effekter ("outcomes of interest"):
- Tiltaket kompenserer for varig markedssvikt i kunnskapsmarkeder => bedriften vil få en bedre løsning på en konkret utfordring
- Tiltaket avlaster risikoen ved FoU-investering => bedriftsporteføljen vil få en bedre økonomisk utvikling
- Tiltaket gir erfaring med ekstern FoU-relasjon og gir mersmak => bedriften vil endre sin FoU-atferd

# Effektstudier som kan brukes

---

- Ikke tenk på evidens som (økonomiske) effektstudier, men som empirisk belegg (kval/kvant) for å forstå hvordan virkemidlene virker
- Tilpass design til hva man trenger for forståelse av problemområde, utvikling av alternative tiltak, kartlegging av ukjente (men viktige) sammenhenger i kausalkjedene – ofte er det viktigere å forstå hva som skjer enn å vite hvor sterk sammenhengene er
- Forbeholdene er ikke bare forbehold, men må leses som en integrert del av resultatene – gjerne med enkle robusthetsanalyser
- Utforsk alternative design og strategier – gjerne i egne forskningsprosjekt, ikke putt alt inn i evalueringsoppdrag



Teknologi for et bedre samfunn