

Å se sannheten i øynene

Hvorfor brukes enorme ressurser på å evaluere tiltak som ikke kan evalueres mens mulighetene for å få vite om noe faktisk virker skusles bort?

Hvert år kjøper det offentlige inn en mengde evalueringsrapporter fra forsknings- og konsulentmiljøer. Det er enighet om at offentlig politikk skal evalueres, være seg ungdomstiltak, kvaliteten på omsorgstilbud, arbeidsmarkedstiltak eller bistand. Problemet er at svært mange offentlige tiltak settes i gang uten en klar idé om hvordan det skal være mulig å i etterkant sjekke om de virker etter hensikten.

Hva er effekter og hvorfor er de så vanskelig å måle?

Effekten av et tiltak er egentlig svaret på et tankeeksperiment der vi tenker oss hvordan det ville gått med den samme personen i én verden der tiltaket settes inn og en annen verden der det ikke settes inn. Dersom det ble vin i butikkene, ville du drikket mer alkohol? Siden de fleste av oss lever kun i et univers og derfor ikke kan oppleve både å få vin i butikken og ikke få vin i butikken på samme tid er det selvfølgelig helt umulig å observere denne effekten direkte. Problemet ligger i å etablere det vi kaller det "kontrafaktiske utfallet", altså det som ikke skjedde. I en randomisert kontrollert studie (RCT) gjør en dette ved å avgjøre hvem som skal få ekte pille og sukkerpille ved loddtrekning. På denne måten sikrer en at disse to gruppene, i forventet forstand, er helt like. Etter at de begge har svelget pillen vil all observert forskjell mellom de to gruppene (f.eks. ulikt blodtrykk, overlevelse m.m., avhengig av preparat) anses som en *effekt* av pillen og enkel statistisk analyse kan benyttes for å sannsynliggjøre om disse forskjellene er kausale eller kun et utslag av tilfeldigheter. Denne typen forsøk kreves for legemidler og det må dokumenteres at de har effekt.

Hva er det så vi gjør når vi skal evaluere andre offentlige tiltak? La meg ta et eksempel, IA-avtalen, som har til hensikt å redusere sykefraværet og samtidig bedre inkluderingen av utsatte grupper i arbeidslivet. Det er blitt brukt mye ressurser på denne avtalen og siden den er en forpliktende avtale mellom regjeringen og partene står den også "i veien" for annen politikktutvikling. Avtalen har klare målsetninger, bl.a en 20 prosent reduksjon i sykefraværet. Alle som kjenner feltet har i lang tid hørt om viktigheten av IA-arbeidet ute på den enkelte arbeidsplass. Ved å bli IA-bedrift skal ens oppfølging av sykmeldte endres. Ledelsen skal følge opp sykmeldte tidlig, etter et fastlagt skjema og samtidig vises de ansatte tillit ved at de gis forlenget periode for egenmeldt sykefravær (fra 3 til 8 dager).¹ 1000-kronersspørsmålet er jo da, virker det? Svaret er dessverre at det er det egentlig ingen som vet. Det finnes ikke en gang gode data for hvilke arbeidssteder, og fra når, som er IA-medlemmer og ikke. Såvidt meg bekjent fantes det en liste hos NAV over IA-medlemmer men denne ble skrevet over hver gang nye kom til slik at det i begynnelsen av avtaleperioden ikke er mulig å vite når de ulike arbeidsstedene ble medlemmer. Dette ble endret slik at datakvaliteten har bedret seg, men for meg sier det ganske mye om mangelen på bevissthet om at tiltak en gang skal evalueres. Selv om data hadde vært av bedre kvalitet ville likevel evaluering fortsatt vært vanskelig. Årsaken er at det på ingen måte er tilfeldig hvilke arbeidssteder som velger å bli IA-medlem og ikke, og det er trolig heller ikke tilfeldig når de velger å inngå medlemskap. Husk at om vi skal tolke alle forskjeller

¹ Det er verdt å si at for mange som jobber med IA-avtalen er denne avtalen nærmest uløselig knyttet til alle politikkkendringer på dette feltet. Fordi det er evaluering av offentlige tiltak, og ikke IA-avtalen, som er tema for denne artikkelen har jeg valgt å legge en snever forståelse av IA-avtalen til grunn, nemlig IA-medlemskap.

på IA-medlemmer og ikke-medlemmer som en effekt av IA-avtalen så må vi tro at de *hadde vært like* dersom ingen eller alle hadde vært IA-medlemmer (etter at vi hadde kontrollert for observerbare kjennetegn). I svært mange tilfeller er det ingen grunn til å tro at dette er tilfelle. Om en virkelig ville ønsket et svar på om IA-arbeidet virket burde en ha randomisert gjennomføringen, f.eks ved å trekke ut et antall arbeidssteder som gjennomførte det eller gjennomført det i tilfeldige kommuner/regioner. Da kunne vi sammenlignet og fått svar.

Hva tror du mest på?

I fjorårets debatt om sykefraværet dukket ofte det såkalte "Tillitsprosjektet" i Mandal kommune opp. Her ble de ansatte vist stor tillit ved at de fikk fri egenmeldingskvote for sykefravær og dette ble kombinert med bl.a. bedret oppfølging av sykmeldte. Det viste seg at sykefraværet ikke gikk i takt men at det i stedet falt – i hvert fall i begynnelsen. Men, var det de frie egenmeldingene, bedret oppfølging eller lokal entusiasme som var årsaken? Et sosial eksperiment fra 1988 i Göteborg og Jämtland kan kanskje gi oss svaret. Svenske myndigheter bestemte da å gi alle født på en partallsdato to uker egenmelding mens alle født på en oddetallsdato fortsatte med én uke. Det viste seg at de som fikk to uker egenmelding også ble lenger borte. Alle arbeidstakerne hadde høyst sannsynlighet for å komme tilbake til jobb dagen før de måtte besøke lege, dvs. dag 7 for dem født på oddetallsdato og dag 14 for dem med partallsdato. Det ble ikke flere sykmeldinger, men de ble lengre, akkurat slik en litt kynisk ville gjettest på forhånd. Tror du tillitsprosjektet eller det sosiale eksperimentet fra Göteborg og Jämtland er best egnet til å si noe om effekten av økt egenmeldingsperiode? Jeg er ikke i tvil.

De gode eksemplers forbannelse

I den politiske debatten hører vi ofte at vi skal "lære av de gode eksemplene". Dette er et blindspor, av to grunner. For det første; for hvert gode eksempel kan det finnes mange som ikke har lyktes med det samme tiltaket, men som vi aldri hører om. For det andre; det er veldig vanskelig å skille ut *hva som virker* fra et godt eksempel. Når *Drillos* lykkes, skyldes det 4-5-1 formasjonen, fokus på raske overganger, dyrking av spissferdigheter eller Egil Olsens evne til å motivere og tenne spillerne – eller er det kanskje hans pedagogiske ferdigheter?

La oss eksperimentere

Skal vi få vite hvordan politikken virker må vi være villige til å gjøre eksperimenter, slik som vi gjør med legemidler. I stedet for å synse om det blir mer eller mindre vold av å stenge alle utestedene samtidig og/eller tidligere enn i dag kan en jo bare teste det ut. Slå terning og la noen kommuner ha fleksibel åpningstid i et år mens andre kommuner stenger tidlig. Fremfor å synse om hvorvidt bistand virker eller ikke kan en lage eksperimenter der en tester ut ulike tiltak, f.eks. i et tilfeldig utvalg av landsbyer (dette gjøres allerede av Poverty Action Lab ved MIT). Fremfor å synse om hvorvidt arbeidsmarkedstiltak virker eller ikke kan en gjøre randomiserte forsøk der grupper av arbeidstakere tilbys ulike tiltak for å finne ut hvilke som virker best.

Om en mener alvor med å bruke offentlige midler på en god måte må en også ta seg tid og bry med å legge til rette for å få svar. Randomiserte eksperimenter er den klart beste måten å finne ut dette på.