



Martin D. Munk

Metoder til at måle kausale effekter af socialpolitiske indsatser

I det politiske system rettes fokus i stigende grad på om socialpolitiske indsatser har en kausal effekt. I artiklen fremstilles krav, der skal være opfyldt til undersøgelsesdesign og metode for, at en undersøgelse efterfølgende kan anvendes til at sige noget om effekten af den indsats, der belyses. Om en indsats har en kausal effekt, eller om der er en årsagssammenhæng mellem indsats og virkning er vanskelig at bestemme og kræver, at et givet udfald kun kan forklares ud fra en indsats og ikke ud fra andre observerede og uobserverede forhold. Klassisk statistik og metoder kan ikke sikre validitet. Økonometriske metoder og strukturelle modeller, som udfoldet i James Heckman traditionen, er i den sammenhæng ikke til at komme uden om, og som særligt har fokus på betydningen af uobserverede forhold, uafhængighedsproblemet, og på om man udgår fra en reel social handlingsmodel, der kan beskrive sammenhængen mellem "input" og "output". Samfundsmæssigt set er det grundlæggende ikke acceptabelt, hvis de socialpolitiske indsatser reelt ikke kan siges at have en kausal effekt.

Søgeord: Kausal effekt, socialpolitiske indsatser, eksperiment, specifikationer, økonometriske metoder, sociale handlingsmodeller.



Formålet med artiklen¹ er at belyse og fremstille krav, der skal være opfyldt til undersøgelsesdesign og metode for at en undersøgelse efterfølgende kan anvendes til at sige noget om effekten af den indsats, der belyses. Der er et stort behov for at styrke designet af danske undersøgelser og evalueringer af den socialpolitiske indsats. Det er relevant at kunne afgøre, om en behandling, en indsats eller en intervention virker. Det kan eksempelvis være en indsats i et dagtilbud til børn. Hjælper det at give børn mere eller bedre sprog-stimulering i børnehaven? Vi ved det faktisk ikke med sikkerhed. Det er naturligvis afgørende at fastlægge, om en indsats virker. Ellers vil vi heller ikke med sikkerhed kunne sige, om børns kompetencer potentielt styrkes gennem forskellige typer af tiltag, eksempelvis med henblik på at børnene bliver mere parate til at gå i skole. Det vigtigt at vide, om disse indsatser bidrager til at udvikle børnenes kompetencer, og om deres kompetencer samlet set bliver bedre.²

Der findes undersøgelsesdesigns og metoder, der potentielt kan anvendes til at afgøre, om en indsats virker. Forskningsområdet, der ligger til grund for disse metoder, er imidlertid under stadig udvikling og korrektion. I forbindelse med anvendelse af sådanne metoder er der en række krav, som skal være opfyldt for at kunne afgøre, om indsatserne har en effekt. For at kunne bestemme, om indsatserne har en *kausal effekt*³ (se næste afsnit om kausalitet) er der en række krav, som skal være opfyldt til designet af de undersøgelser, der iværksættes, og til de enkelte metoder, som efterfølgende anvendes til at bestemme eller finde effekter. Der findes en betydelig litteratur inden for området, men den er samtidigt også ganske heterogen og modstridende.

Et overordnet formål med effektstudier er, at brugen af indsatser over for fx børn gøres målbar og sammenlignelig med henblik på systematisering af en samlet indsats. Det er således hensigten at skabe grundlag for brug af systematiske effektmålinger, som gør det muligt at måle effekter på såvel kort som på lang sigt. Forsøgsdesign skal derfor sikre muligheden for måling af kortsigtede og langsigtede effekter. I artiklen beskrives de grundlæggende elementer og krav-specifikationer til design og metode, og artiklen bygger hovedsageligt på den statistiske litteratur om emnet, men det er vigtigt at understrege, at der er problemer i denne litteratur, som blandt andre økonometrikeren James Heckman har påpeget. De præsenterede begreber har indbyggede svagheder, som ikke i fuldt omfang problematiseres her.

Kausalitet

Kausalitet er et omdiskuteret begreb, og gennem tiden har der i videnskabelige kredse været en intens og stridende debat om, hvad kausalitet dækker over (se fx Woodward 2003, for en generel filosofisk diskussion af kausalitet). Der er imidlertid næppe tvivl om, at viden, der er legitimeret som kausalt funderet, ofte tillægges større vægt. Økonometrikeren Heckman (2005a:1) definerer ud fra primært John Stuart Mill følgende:



Martin D. Munk
Fil.dr. i sociologi,
seniorforsker ved SFI
– Det Nationale
Forskningcenter for
Velfærd
E-mail: mdm@sfi.dk

Kausalitet er et meget intuitivt begreb, som det er meget vanskeligt at gøre præcist uden at gøre det tautologisk. To ingredienser er centrale for enhver definition: (1) et sæt af mulige udfald (kontrafaktiske) genereres af et sæt af "faktorer" eller "determinanter" og (2) en manipulation, hvor en (eller flere) "faktorer" eller "determinanter" ændres. En effekt realiseres som en ændring af en stabil funktion, der giver en ændring i udfaldet baseret på et sæt af interventioner som foranlediger, at "faktorerne" ændres med samme størrelse (egen oversættelse).⁴

I metodelitteraturen findes der forskellige kausal begreber, herunder deterministisk kausalitet, kausalitet baseret på et sandsynlighedsbegreb, økonometrisk kausalitet, kausalitet baseret på eksperimenter og kausalitet i kausale modeller (jf. Sobel 1995, se også Winship og Sobel 2004, for en lettere tilgang til de forskellige begreber og modeller). Det er især kausalitet baseret på eksperimenter, kontrafaktisk (Donald Rubin) tænkning og økonometrisk fundering, som artiklen bygger på. Det kan naturligvis diskuteres, om beskrivelse kan dække et socialt fænomen, eller om det altid er nødvendigt at kræve kausale forklaringer på et givet socialt fænomen, eksempelvis i forhold til større historiske begivenheder.

For at gøre diskussionen om kausalitet lidt mere nærværende trækker jeg i det følgende på et illustrativt eksempel, som Morgan og Winship (2007) anvender i deres bog om kausale modeller. Den velkendte Wisconsin-modellen, som er designet til at kunne forklare status niveau (status attainment), postulerer, at studerendes uddannelsesniveau bl.a. bestemmes af egne forventninger eller aspirationer. Foruden familiebaggrund og mentale evner spiller andres forventninger til de studerende samtidigt ind på opnået uddannelsesniveau. Forventningerne udspringer fra signifikante andre, som navnlig tæller forældre, lærere og kammerater. Disse forventninger internaliseres efterhånden og bliver til de studerendes egne forventninger, der udgør grundlaget for motivationen til at opnå en bestemt uddannelse. Denne model indebærer – kausalt – at de studerendes forventninger kan øges ved, at de signifikante andre øger deres forventninger til dem. Denne tankegang er imidlertid blevet kritiseret af Pierre Bourdieu (1973) for ikke at tage højde for strukturelle be-



grænsninger i de objektive muligheder. Med andre ord, der er i realiteten ikke etableret en kausal sammenhæng mellem aspirationer og uddannelsesniveau, da det er de objektive muligheder, som er ulige struktureret, der bestemmer de studerendes mulige uddannelsesniveau, og derfor de facto betinger, i hvilken grad forventningerne kan indfries.

Indsats og kravspecifikationer

I det følgende beskrives indsatsbegrebet. Desuden beskrives de overordnede kravspecifikationer til fremtidige undersøgelser af indsats, som kan tages i anvendelse med henblik på at optimere designet. I forbindelse med iværksættelse af undersøgelser af indsats og effekter belyses indsats, der kan anvendes, og om en given indsats eller intervention virker.

Begrebet indsats anvendes om forskellige former for indsats, behandling, intervention og tiltag, som typisk anvendes eller afprøves på en eller flere grupper eller enheder. De omhandler således især måling af konkrete effekter, som beskrevet i Jespersen (2006), men vil også med fordel kunne anvendes i forbindelse med undersøgelser af social ulighed og marginalisering. En indsats kan eksempelvis være brug af samlede curricula, øget sprog-stimulering, brug af specifikke kropstræningsprogrammer eller mere generelt dagpasning af høj kvalitet.

En amerikansk barndomsforsker Janet Currie sammenfatter i bogen *The Invisible Safety net* (2006, især kapitel 5) betydningen af høj kvalitet i pasningen af småbørn og gennemfører en vidtgående analyse af de amerikanske erfaringer. Dagpasning af høj kvalitet kan bidrage til at mindske den sociale reproduktion⁵, hvilket bekræftes af undersøgelser i udlandet. Mere specifikt kan denne indsats være at øge fokus på en førskoleorienteret læring med henblik på at styrke skoleparathed, dvs. evnen til at indgå i en læringssituation i de første klasser i skolen. Denne evne er sammensat af forskellige elementer som *kognition*, sociale kompetencer, selvdisciplin, adfærdsregulering, glæde og indlevelsessevne. En styrkelse af denne evne kan føre til umiddelbart positive effekter ved skolestart, herunder forbedrede matematik- og læsekompetencer, men disse effekter viser sig i gennemsnit at falde bort knapt et år efter skolestart. Desuden bidrager tidlig førskole målrettet læring til adfærdsproblemer blandt børnene. Imidlertid tyder det på, at tidlig førskole læring kan give dårligt stillede børn en ekstra løftestang, skolemæssigt set (jf. Magnuson, Ruhm & Waldfogel 2007).

I den internationale forskning på området er der ikke fuldstændig enighed om, hvad skoleparathed dækker over, idet der dog er en vis enighed om de elementer, der er nævnt ovenfor (jf. Magnuson, Meyers, Ruhm & Waldfogel 2004). James Heckman (se under litteratur) har i flere publikationer peget på, at såvel kognitive som *ikke-kognitive evner/færdigheder* er afgørende for indlæringen og dermed skoleparathed, samt et senere positivt udfald i uddannelsessystemet. Dette bekræftes også i et dansk studie af uddannelseschancer, hvor McIntosh & Munk (2007a) finder, at en motivationsfaktor som glæden ved at



gå i skole har en selvstændig positiv effekt på senere uddannelsesniveau. Det er imidlertid til stadighed en udfordring at afdække, hvad ikke-kognitive evner/færdigheder dækker over, og det er vanskeligt at pege på, hvordan man skal operationalisere disse (jf. Heckman & Rubinstein 2001; Heckman 2000; og Bourdieu 2000, for en indkredsning af ikke-kognitive dispositioner). Studier viser, at en manglende inddragelse af aspekter som motivation og ikke-kognitive færdigheder indebærer en undervurdering af betydningen af at investere i de tidlige indsats i børns liv.⁶

Uenigheden i den internationale forskning understreger vigtigheden af at anvende en række kravspecifikationer, og at man til stadighed arbejder med design og metode i forbindelse med iværksættelse af nye undersøgelser, dels for at kunne vurdere tidligere anvendte begreber og resultater og dels med henblik på at vælge korrekte og relevante metoder og begreber. På den baggrund er det relevant at stille spørgsmålet: Hvilke krav til design og metode skal være opfyldt for, at man med sikkerhed kan afgøre, om en indsats har en positiv eller negativ effekt i forhold til et ønsket resultat som fx skoleparathed, præstation i en faglig test i skolen, senere uddannelsesniveau eller mindsning af marginalisering og social ulighed? Med andre ord, hvilke krav skal man stille til et undersøgelsesdesign for at sikre, at effekter kan siges at være evidensbaseret og samtidig ikke er *biased*?

En række designs og metoder er at foretrække, når effekten af en indsats skal undersøges. I USA har forskere i løbet af de sidste 40 år gennemført undersøgelser ved hjælp af designs og metoder, der lever op til eller forsøger at leve op til en række kravspecifikationer. En del af disse undersøgelser – fx Perry Preschool studiet – har en betydelig værdi, dels policymæssigt, dels forskningsmæssigt (se fx Currie & Thomas 2000; Currie 2001; Garces, Currie & Thomas 2002; Magnuson & Waldfogel 2005). Undersøgelserne har endvidere bidraget til at rette fokus på indsatser, der dels virker på kort sigt, dvs. graden af skoleparathed, og dels på længere sigt, dvs. på senere udfald som eksempelvis uddannelsesniveau, deltagelsesgrad på arbejdsmarkedet og indkomst. Disse studier har bidraget til, at amerikansk forskning i dag indtager en førende position på området, se fx artikler i *Handbook of Economics of Education*, redigeret af Eric Hanushek & Finis Welch (2006).

Det afgørende er, at de effekter man måler, er kausale, dvs., at de udtrykker en reel årsagssammenhæng mellem indsats og virkning. Dette kan opfyldes ved at designe undersøgelsen på en korrekt måde. Et 100 pct. tilfredsstillende design opnås typisk ved at gennemføre et socialt eksperiment, men det kan i praksis være vanskeligt at gennemføre, især på grund af etiske og politiske aspekter og dilemmaer. Fx er det svært at forestille sig, at man kun ville øge andelen af uddannede pædagoger i udvalgte daginstitutioner inden for en kommune. Det ville indebære, at andre daginstitutioner i en kommune ikke ville få dette tilbud med deraf følgende potentiel utilfredshed fra forældrene i de daginstitutioner, der ikke var omfattet.



Ekspérimentdesignet (det sociale eksperiment)

Et krav til undersøgelser er, at de frembringer *signifikante* og *unbiased* resultater. Disse krav kan opnås ved at gennemføre et klassisk socialt eksperiment, hvor man tilfældigt udvælger en indsatsgruppe og en kontrolgruppe. Dette kaldes for *randomiseret kontrolleret forsøg* (Randomised Control Trial eller RCT; se fx Macdonald 2001, eller Cook & DeMets 2008) eller for det sociale eksperiment (jf. Manski 1995; Greenberg & Morris 2005). RCT indebærer, at en gruppe omfattes af eller gives en indsats, medens kontrolgruppen ikke omfattes af eller gives en indsats. Udvælgelsen foretages på en sådan måde, at indsatsgruppen og kontrolgruppen er tilfældigt (randomiseret) udvalgt, og baggrundsfaktorerne for de to grupper er lig hinanden. Derved er det muligt at sammenligne de to grupper, efter at indsatsen er givet til den gruppe, der er omfattet af indsatsen. Den forskel, der eventuelt efterfølgende viser sig mellem de to grupper kaldes for *treatment effekten* (forskellen mellem de behandlede og ikke-behandlede opnås fx ved at udregne differencen mellem gennemsnittet for hver af de to grupper). Dette design betjener sig af det klassiske "assigned treatment setup", hvilket indebærer, at en gruppe (fx en stikprøve af personer) med givne observerede baggrundsforhold tilfældigt bliver udvalgt til behandling. Alle personer får den samme behandling. Dette design blev allerede i 1935 beskrevet i *The Design of Experiments* af den berømte statistiker R. A. Fisher (jf. Manski 1995; Campell & Stanley 1963/1966; Shadish, Cook & Campell 2002).

Angrist (2004, 2001) peger på, at det randomiserede forsøg ikke er den eneste måde at designe et studium på, men det er den type af design, som sætter standarden for andre studier. Det sociale eksperiment er langt det bedste som evalueringsstrategi, idet designet giver mulighed for at eliminere forstyrrende indflydelse fra uobserverede forhold. I sidste ende bliver det dog en afvejning af fordele og ulemper, der betinger valget af design.

Quasi-eksperimenter (naturlige ikke randomiserede forsøg)

Det randomiserede forsøg (RCT) er af forskellige grunde ikke altid muligt at gennemføre, fx kan det som tidligere nævnt være vanskeligt at forestille sig, at et tilbud/en indsats kun gives nogle udvalgte daginstitutioner i en kommune, mens de andre daginstitutioner i kommunen ikke får tilbuddet. Et andet design kan imidlertid også anvendes, nemlig det såkaldte quasi-eksperimentelle design, eller det naturlige eksperiment (kaldes for natural experiment i metode-litteraturen), som også den tidlige sociologi opdyrkede, eksempelvis Samuel A. Stouffer og F. Stuart Chapin. Sidstnævnte blev i 1940'erne kritiseret for sin bog *Experimental Designs in Sociological Research*, men Chapins fremhævelse af matchning og quasi-eksperimenter blev senere fulgt op af Campell & Stanley (jf. Morgan & Winship 2007:6-9).

Et naturligt eksperiment defineres som et observationstudie, der fra naturens hånd er randomiseret, dvs. har en tilfældig variation (Freedman 2005:6).



Et naturligt ikke-randomiseret design er således attraktivt, hvis man kan finde en kilde med en naturlig variation, som fx hvilke daginstitutioner der anvender sprogstimulering, og hvilke der ikke gør det, både inden for en kommune og mellem kommuner. Det naturlige eksperiment kan således erstatte et rigtigt eksperiment ved at finde en gruppe, som allerede "naturligt" får behandlingen eller en given indsats, og en tilfældig kontrolgruppe som ikke får behandlingen. Derfor er et quasi-eksperiment et naturligt eksperiment, som under alle omstændigheder foregår ved, at en gruppe af børn påvirkes via en eksogen begivenhed⁷, medens en anden gruppe af børn, som ikke er påvirket af denne begivenhed, fungerer som kontrolgruppe. Den eksogene begivenhed kan være et naturligt forekommende fænomen, en specifik politik eller pædagogik, der føres i givne institutioner. Disse naturligt forekomne fænomener kan tænkes at påvirke adfærden hos en gruppe børn. Det kunne eksempelvis være tidlig sprogstimulering eller specifikke former for leg, som gennemføres i udvalgte daginstitutioner i en kommune. Man kan også tage udgangspunkt i forskelle mellem kommuner som eksempelvis normering, pasningsgarantier, personalepolitik og uddannelsespolitik. Det er dog afgørende, at det er de facto iværksatte tiltag og ikke blot er planlagte tiltag. Ellers er det ikke muligt at måle en kausal effekt.

Når man ønsker at sammenholde en behandlet gruppe og en kontrolgruppe, er det oplagt at anvende en matchning metode, hvor man "parrer" gruppen, der behandles med en indsats med en gruppe, der karakteristisk ligner gruppen, men som ikke behandles. Denne situation med matchning kan eksempelvis etableres ved at sammenligne daginstitutioner i en kommune, som gennemfører en indsats, med daginstitutioner i andre kommuner, som ikke gennemfører indsatsen. Kravet er, at de givne daginstitutioner er sammenlignelige for så vidt angår observerede karakteristika (rekruttering af personale, social sammensætning, lokale forhold etc.). Ved matchning *antages* det, at når der kontrolleres for observerbare karakteristika, så er kontrol- og behandlingsgruppe også ens på uobserverbare karakteristika, og derved tages der i princippet højde for uobserverede forhold. Denne antagelse er dog blevet kritiseret af Heckman (2005a, 2005b). Et quasi-eksperiment kan etableres ved at benytte tilgængelige data, eksempelvis kommunedata, registerdata eller surveydata med en naturlig variation, hvorved en gruppe via en specifik indsats og baggrundskarakteristika kan udvælges. På tilsvarende vis udvælges en kontrolgruppe, som ikke får behandlingen, men som på baggrundskarakteristika matcher den behandlede gruppe. Ideen er, at den sidste gruppe ikke påvirkes af indsatsen, men er sammenlignelig med den behandlede gruppe.

Det afgørende for at kunne opstille et korrekt quasi-eksperiment er, at det er klart defineret, hvad der identificerer forskellen mellem at blive behandlet og ikke at blive behandlet, eller med andre ord at være i indsatsgruppen eller i kontrolgruppen. En anden forudsætning er, at individerne, som findes i henholdsvis den behandlede gruppe og i kontrolgruppen, ligner hinanden. Hvis



ikke denne forudsætning er opfyldt, kan der på forhånd være systematiske forskelle mellem børnene i de to grupper, og dermed opstår det såkaldte *selektionsproblem* (bias), idet nogle børn i den ene gruppe måske allerede har en fordel sammenlignet med den anden gruppe af børn. Det kunne fx være, at børnene får en anden støtte i af form midler til økologisk frugt, ekstra midler til ture, bedre fysiske udfoldelsesmuligheder eller stor forskel i erfaringskapitalen, som pædagogerne besidder. Er disse to forudsætninger opfyldt, er det muligt at kunne beregne en kausal effekt af indsatsen for de, der behandles. For at hindre selektionsproblemer må der således hverken være forskelle i observerbare eller uobserverbare karakteristika, der udløser en forskel mellem de behandlede, og de som tilhører kontrolgruppen (Heckman m.fl.).

Et problem kan være, at en af grupperne, der sammenlignes før og efter behandling, undervejs ændrer karakter enten på gruppeniveau eller på det enkelte individplan (jf. Winship & Morgan 1999). Et andet problem man skal være opmærksom på er, hvis den naturlige variation er endogen, dvs. afhængig eller påvirket af at være en specifik gruppe. Der kan fx være tale om en ressourcestærk gruppe af forældre, som er i stand til at få en bestemt ordning gennemført i deres institution for deres børn og derfor umiddelbart ser ud som en naturlig variation, men i realiteten ikke er det.

Med andre ord er det afgørende, at *alle* observerbare og uobserverbare karakteristika er ens for de to grupper, der sammenlignes. Kravet om, at observerbare karakteristika skal være eksogene kan variere. I Cobb-Clark & Crossley (2003) refereres til studier, der dels opfylder kravet om, at de observerbare variable skal være eksogene, dels til studier hvor det modsatte er tilfældet.

Konklusionen er, at kontrol- og behandlingsgruppen skal være ens på *såvel* observerbare som uobserverbare karakteristika, hvilket sikres gennem det sociale eksperiment eller det naturlige eksperiment.

Ikke-eksperimentelle designs i øvrigt

En tredje mulighed vil være at anvende andre observationsdata, hvor der er rig adgang til både registre og surveys. Hvis en sådan metode anvendes, er kravet, at der tages højde for en række forhold svarende til dem, der er nævnt under den quasi-eksperimentelle metode. Denne metode er dog klart mindre optimal end de to andre, hvor man langt lettere kan leve op til kravene.

Definition af indsats og effekt i statistisk forstand

Statistisk set er formålet at kunne beregne (estimere) en gennemsnitlig treatment (indsats) effekt (*ATE*), som er kausal eller mere interessant en gennemsnitlig effekt på betingelse af at have været behandlet (*ATET*), dvs. typisk en binær forklarende variabel, med andre ord en ja/nej variabel. I det følgende trækkes på fremstillingen i Wooldridge (2002), Blundell, Dearden & Sianesi (2004, 2005), Angrist (2004) samt Winship & Morgan (1999), se også Morgan &



Winship (2007). Estimering af ATE står helt centralt i programevalueringslitteraturen. Tricket er at opstille et kontrafaktisk system (counterfactual system), fremsat af Rubin i 1974 i et tidsskrift for pædagogisk psykologi. Dette trick blev senere overtaget af en række forskere inden for samfundsvidenskab og statistik, blandt andre førende forskere som Imbens og Angrist. Dette system tillader os at definere forskellige former for indsatser, som har en policymæssig interesse.

Lad y_1 betegne et udfald efter behandling (treatment) og y_0 betegne et udfald uden behandling. Det antages, at et individ, eller individerne, ikke optræder i begge tilstande. Vi har indtil videre ikke nogen antagelser om fordelingerne i de to grupper. De kan være kontinuert fordelte, hvis det var indkomster, men ofte er det netop binære udfald, dvs. to mulige udfald. Eksempelvis blev løftet af øget sprogtræning i dagtilbud, blev ikke løftet af sprogtræning i et dagtilbud. I det følgende antages, at vi har uafhængige identiske fordelinger for de to udvalgte stikprøver/grupper. Denne antagelse gælder ikke, når behandlingen har en effekt på et andet udfald, eksempelvis stigmatisering af en gruppe af børn, som ikke reagerer negativt på sprogtræning, eller at en ændring i behandlingsstatus for ethvert individ ikke påvirker de potentielle udfald for de andre individer. Antagelsen at en behandling kun påvirker et udfald, kaldes for SUTVA (stable unit treatment value assumption) antagelsen (Wooldridge 2002:604). Det antages således, at virkningen af en indsats eller intervention er konstant for et enkelt individ. Denne antagelse holder imidlertid ikke, når forskellige indsatser interagerer med hinanden (Winship & Morgan 1999:663; Heckman 2005a:38; Heckman 2005b:138), eller hvis indsatsen ikke er konstant for det enkelte individ og varierer over tid. Den størrelse, man således er interesseret i, er den gennemsnitlige behandlingseffekt af de behandlede. Denne defineres som: $ATE_1 = E(y_1 - y_0 | w = 1)$ (Average Treatment Effect on the Treated = ATET), hvilket er den gennemsnitlige behandling for de individer, som rent faktisk behandles. Det er således gennemsnittet for dem, der er blevet behandlet; $w=1$ angiver betingelsen for, at behandlingen er foregået. (E betegner forventningsoperatoren, som er en formel måde at skrive et gennemsnit på; den gængse ikke-matematiske definition er populationsgennemsnittet i en uendelig stor population). Det er med andre ord den forventede effekt af, at en person er behandlet eller har deltaget i et program (Angrist 2004 terminologi), som er udtrukket tilfældigt fra en population.

Det er relevant at gøre opmærksom på, at Heckman (2005a:6) har kritiseret den kontrafaktiske model (som han kalder for den hypotetiske model); han har i stedet foreslået strukturelle økonometriske modeller, som indebærer analyse af såvel objektive udfald som subjektive vurderinger.⁸ En holdbar analyse fordrer, at der defineres klare modeller for udfald og indsatsvalg byggede på teorier om menneskelig social handling og adfærd, hvor de uobserverbare forhold, som ligger bag selektions- og evalueringsproblemet, gøres eksplícit (jf. Heckman 2005b). (Se videre under økonometriske metoder).



Kausalitetsmodeller i statistik er således ufuldstændige, da de hverken specificerer mekanismer fra den eksterne variation eller tilfældighed, som begge er med til at bestemme udfaldene, og relationen mellem disse og selektionsmekanismerne (Heckman 2005a:85).

I den forbindelse er det relevant at være opmærksom på de af Heckman opstillede tre policy-evaluerings-problemer, fordi de styrker fokus på kausalitet i relation til konkrete problemer: 1) Har historiske interventioner, dvs. faktiske erfarede indsatser, en indvirkning på forskellige udfald, fx på velfærd? Dette problem omhandler således intern validitet, dvs. identifikation af den enkelte indsatsparameter, 2) Kan man forudsige, om interventioner udviklet i et system har nogen som helst virkning i andre systemer? Dette problem omhandler, hvad man typisk kalder for ekstern validitet, 3) Kan man forudsige virkningerne af nye interventioner (nye politikker), som aldrig er anvendt, givet at der er opstillet et kontrafaktisk system? Dette problem forudsætter, at man anvender tidligere tiders erfaringer med henblik på at forudsige konsekvenser af nye (politikker). I Heckman (2005a) omtales de anbefalede metoder, der kan tages i anvendelse, men de vil ikke i detaljer blive beskrevet her.

Statistiske korrektioner

I forbindelse med anvendelse af statistiske metoder er det et krav, at der tages højde for sædvanlige statistiske problemer som selektion, uobserverede heterogenitet/faktorer, heteroskedasticitet (varierende varians) og compliance. Selektion kan fx være, at der mangler eller er meget få svage børn i den udvalgte gruppe, som får behandlingen, fordi indsatsen måske realiseres i områder med en vis overskudskapacitet af den ene eller anden art. Dette kan gøre det svært at undersøge, om en indsats har en virkning, da indsatsen ikke nødvendigvis anvendes på den helt rigtige gruppe. Et argument i metodelitteraturen er, at det kan være tilstrækkeligt at få den gennemsnitlige virkning frem i en gruppe, og at man ikke nødvendigvis opnår, at alle involverede i gruppen er blevet påvirket af indsatsen, eller med andre ord, at en effekt virker på alle involverede (Winship & Morgan 1999).

Det er dog et erkendt selektionsproblem, at de observerede i gruppen, som har fået behandlingen, kan have fået denne, fordi man selv-vælger sig ind på et bestemt program eller en institution, som forældrene til børnene eksempelvis vælger sig ind på. Det bliver et problem, når estimeringen skal foretages. Selektionsproblemet består således i, at personer selv vælger eller bliver valgt ind i behandlingen på baggrund af *uobserverbare* karakteristika. Et andet problem, som også kan opstå, handler om, at behandlingen kan virke forskelligt på de involverede. Fx kan det tænkes, at indsatser målrettet til udsatte børn kunne komme til at virke stigmatiserende. Det kunne indebære, at selve indsatsen faktisk har en negativ utilsigtet effekt. Problemet med heterogene behandlingseffekter er, at det muligvis er et stort kunststykke at ekstrapolere fra den gruppe, som har modtaget behandling, til *hele populationen*, hvis de



behandlede slet ikke minder om resten af populationen.

Uobserverede (uoplyste) forhold kan føre til, at betydningen af en given indsats (af de observerede forhold) overvurderes eller undervurderes, fordi der er forhold, man har dårlige oplysninger om eller slet ikke har oplysninger om. Det statistiske problem er, at de uobserverede forhold kan hænge sammen med de observerede karakteristika. Oplysninger om forældrene til børnene er afgørende, men typisk findes ikke alle oplysninger. Det kan være, at man har oplysninger om forældrenes uddannelse og arbejde, men ikke oplysninger om deres kapacitet i øvrigt, forældrerolle, attituder, adfærd, evner eller dispositioner. Det er heller ikke sikkert, at man har standardiserede oplysninger om børnenes evner eller dispositioner. Dette kan der tages højde for i forbindelse med brugen af de statistiske metoder på en sådan måde, at det der ikke foreligger oplysninger om, ikke påvirker målingen af effekten af indsatsen.

Endelig er det i forskellige studier ofte nødvendigt at tage højde for endogenitet, idet de forklarende variable ikke altid er eksogene (uafhængige). De er i virkeligheden selv udfaldsvariable (afhængige variable).

Det er under alle omstændigheder vigtigt, at de daginstitutioner, der kan siges at blive behandlet eller anvender en indsats, rent faktisk gennemfører denne for alle involverede. Ellers kan der opstå et *compliance* problem, dvs. der kan være personale eller børn, som ikke følger indsatsen, enten helt eller delvist. Det er et problem, man især har kendskab til inden for den medicinske forskning. Det er afgørende at der tages højde for, at der kan være personer, som ikke indvilger i behandlingen eller indsatsen. Endvidere er det yderst vigtigt, at de der er i kontrolgruppen ikke modtager behandlingen.

Økonometriske metoder

I den økonometriske litteratur findes en del statistiske modeller til at analysere data for at tage højde for problemerne beskrevet i det foregående afsnit om statistiske korrektioner. En af metoderne (matchning) anvendes især med henblik på at kunne sammenligne behandlede grupper med ikke-behandlede grupper (kontrolgruppen). Modellerne omtales oftest som "treatment effects" modeller (jf. Goldberger 1972; Maddala 1983; Greene 2003) og omfatter metoder som regression, matchning og instrument variabel procedure. Den typiske statistiske metode er den velkendte regression, hvor man lineært forsøger at forklare et udfald via en række baggrundsvariable. Derudover eksisterer en familie af tilgange til at modellere treatment effekter, og omfatter:

– Heckman-type' selektionsmodeller (sample selektionsmodel), hvor man samtidigt regner på effekten af en behandling og på selektion i data, som på en eller anden måde er sorteret fra. Desuden antages det, at fejleddene i de to samtidige beregninger er normalfordelte (jf. Buckley & Shang 2003). I Heckman (2005a, 2005b) understreges endvidere betydningen af at arbejde med en egentlig strukturel model, hvor det kan antages, at forskellige outcomes er sammenhængende.



– Beregningsmetoder baseret på at have gode instrument variable til rådighed, hvor en variabel kan defineres i forhold til den selekterede behandling, men som ikke er relateret til de uobserverede forhold og til det målte udfald. Et instrument kan sikre, at man kan generalisere resultaterne. Denne metode uddybes i et efterfølgende afsnit.

– Endelig anvendes *ikke-parametriske* matchning metoder, dvs. metoder, hvor man ikke har strenge antagelser om forudsætningerne/fordelingerne (Green 2003:425). Her udnyttes, at behandlede og ikke-behandlede observationer kan have den samme fordeling for regressorerne (her anvendes den såkaldte propensity score, som vi ikke kommer nærmere ind på), dvs. de forklarende variable (jf. Smith 1997; Winship & Morgan 1999; Wooldridge 2002; Blundell et al. 2004, 2005; Morgan & Harding 2006).

Matchning og regression er forskellige metoder, men foretager ultimativt den samme ting, der går ud på at sammenligne udfald, der er forskellige for så vidt angår indsatsen eller behandlingen, men som er ens på alle baggrunds-karakteristika. Som tidligere nævnt antages det ved matchning, at når der kontrolleres for observerbare karakteristika, så er kontrol- og behandlings-gruppe også ens på uobserverbare karakteristika. Regression bygger som nævnt på linearitet i forhold til udfaldsvariablen, mens matchning potentielt er en ikke-parametrisk procedure, dvs. det er ikke nødvendigt at operere med strenge antagelser om fordelingerne i de to sammenlignede grupper, som anvendes til at identificere behandlingseffekten på udfaldet. Det kan betyde, at det kan være vanskeligt at konkludere på sikker grund. Imidlertid har Heckman og andre opstillet nødvendige forslag til, hvordan en matchning estimator kan blive optimal, som imidlertid ikke er tilstrækkelige (jf. Cobb-Clark & Crossley 2003:509). For det første er det afgørende, at de behandlede og de ikke-behandlede udtrækkes på samme måde. For det andet skal de behandlede og ikke-behandlede observationer kunne sondres fra undersøgelsesinstrumenterne, således at de fundne behandlingseffekter ikke blandes sammen med disse instrumenter. For det tredje skal de behandlede og de ikke-behandlede udtrækkes fra det samme relevante geografiske område, hvilket kræver gode geografiske oplysninger. Endvidere er paneldata rigtig gode at have, fordi man kan foretage før- og eftermålinger for både de behandlede og ikke-behandlede grupper.

Specifikke estimeringsmetoder (IV og difference-in-difference)

Instrument variabel estimation (IV) er parallel med regressionsmetoden. IV metoden anvendes til at håndtere problemer med, at de forklarende variable er afhængige af andre variable, som ikke er observeret, herunder fejleddet i en regressionsligning. Et instrument er således en forklarende eksogen variabel, der har til formål at virke forklarende, og som dæmper eller erstatter en anden variabel, som anvendes til at forklare udfaldet. Dvs. at den nye variabel (instrumentet) skal netop forklare eller hænge sammen med den kausale forkla-



rende variabel, som man i første omgang har til rådighed, men den må ikke hænge sammen med fejleddet eller med andre uobserverede variable, og den må ikke forklare udfaldet. Løses problemet ikke, vil estimatoren ikke være konsistent (unbiased). Problemet er således, at betydningen af den endogene variabel overvurderes, hvis man ikke har et instrument. Man eftersøger derfor et instrument for behandlingen, der minder om den endogene variabel, uden at den – i princippet – er med til at forklare udfaldet (undtagen gennem den kausale variabel, jf. Morgan & Winship 2007:197). IV metoden kræver et vist antal observationer, for at parameterestimatet bliver konsistent.

Det kan ske, at nogle individer i kontrolgruppen lykkes med at få behandlingen. I det tilfælde kan LATE (Local Average Treatment Effect) udregnes, dvs. den gennemsnitlige effekt for de personer, der undervejs i forsøget er kommet i behandling. Det gøres ved at beregne forskellen mellem gennemsnittet for henholdsvis de behandlede og de ikke-behandlede og sandsynligheden for at blive behandlet betinget af, at man får tilbuddet om behandling, minus sandsynligheden for behandling på betingelse af, at man var udtrukket til kontrolgruppen (denne metode kan dog være vanskelig at anvende, jf. Todd 2001 i Angrist 2001).

Når forskellen mellem den behandlede gruppe og kontrolgruppen måles, er det typisk at foretage en simpel sammenligning af den samme gruppe før og efter behandling, hvor sammenligning af et gennemsnit kan kaldes en difference estimering. Er et paneldatasæt til rådighed med observationer for de to grupper over tid, kan en difference-in-difference estimering anvendes til at vurdere effekten af indsatsen, se Abadie (2005). Disse estimators er en slags IV-estimatorer, idet de gennemsnit der beregnes kan opfattes som instrumenter. En anden tilsvarende metode er "regression discontinuity design" (Wooldridge 2002). Her bruges en tærskelværdi for, hvornår man udvælges til behandling. Det kan fx være et handicap, aldersgrænse eller forældresammensætning. En estimering af effekten foretages ved at beregne gennemsnittet for de, der er under tærskelværdien og over tærskelværdien, og som dermed sammenlignes.

Afslutning

Fremover bliver det relevant at kunne afgøre, om en behandling, en indsats eller en intervention virker. Ved at anvende enten sociale eksperimenter eller mere sandsynligt naturlige eksperimenter er det muligt at kunne isolere kausale effekter af en behandling. Vi har i Danmark adgang til rige data som registre og surveys, hvilket giver store muligheder for at opstille naturlige eksperimenter. Forudsætningen for at kunne finde frem til kausale effekter er at anvende økonometriske metoder og strukturelle modeller, der inden for de seneste 10-15 år er blevet kraftigt udviklet. Disse metoder kan give gode muligheder for at undersøge effekten af indsatser og behandlinger i det socialpolitiske system. Lever metodeanvendelsen ikke op til krav som gode data, påli-



delige beregninger (biasfri estimater) og løsning af selektionsproblemer, kan det ikke forventes, at resultaterne er gyldige og generaliserbare.

Etablering af og opbygning af en pålidelig og mere sikker viden vil kunne bidrage til, at man i stigende grad bliver i stand til at forklare, hvilke indsatser der virker, og hvilke som ikke virker. Der kan blive tale om muligheder for at vurdere positive og negative effekter inden for en lang række områder, der er vigtige i den socialpolitiske indsats. Indsatser med det formål at øge børns og unges kompetencer er i den forbindelse et af de vigtigste områder at sikre mere sikker viden om.

Ordforklaringsliste (anvendte teoretiske-statistiske termer)

ATE: Vi kan ikke forvente at kunne observere forskellen $y_1 - y_0$, men vi kan derimod forvente at estimere den gennemsnitlige forskel mellem disse to udfald. Denne defineres som følger: $ATE = E(y_1 - y_0)$ (Average Treatment Effect). En række forskere har betegnet denne situation som mindre relevant, fordi den beregner gennemsnittet for hele befolkningen og dermed inkluderer personer, som egentlig ikke skulle have behandlingen. Det er en kritik, som især James Heckman har fremført via et eksempel, nemlig at det ikke giver mening at gennemføre et jobtræningsprogram for alle, fx er programmet ikke relevant for millionærer, men kun for de grupper det er relevant at gennemføre programmet for (jf. Winship og Morgan 1999:666; Wooldridge 2002:604). Derfor er det mere relevant at gennemføre en indsats eller et program for en mere selekteret gruppe. Det er på den baggrund, at ATET defineres.

Baggrundsfaktorer eller forklarende faktorer: køn, alder, uddannelse, indkomst, boligforhold.

Biased: Fejlbehæftet; eksempelvis fejlbehæftede estimater/koefficienter (det modsatte af unbiased).

Eksogen: Uafhængig – refererer til det modsatte af endogen (se dette).

Endogen: Afhængig variabel. Endogenitetsproblemer får man, når der betinges på en variabel, der er afhængig. Det er statistisk set ikke acceptabelt. Hvis man vil forklare indkomst ud fra uddannelsesniveau, hvilket er en rimelig ambition, så kan uddannelsens betydning overvurderes såfremt man ikke tager højde for familiebaggrund og evner mv., fordi disse baggrundsforhold er med til at forklare uddannelsesniveaut.

Ikke-kognitive evner dækker over ambitioner, pålidelighed, selvdisciplin, organisatoriske evner etc. og refererer således til kompetencer og evner, som er med til at forklare, om et barn klarer sig godt. Såvel kognitive evner som ikke-kognitive evner er vigtige.



Kausal: Kommer af causa (grund) og betyder årsagsbestemt, dvs. en årsag altid fremkalder samme virkning. Man taler om, at der kan være en årsags-sammenhæng mellem to eller flere variable. Kausal effekt indebærer, at en årsag (variable) kan forklare et udfald, når man har kontrolleret for alt andet, såvel observeret som uobserveret. Det kunne fx være, at variabelen familie-baggrund kausalt kan forklare sprogniveau eller evnen til at løse faglige opgaver. Eller et andet eksempel kunne være, at sprogstimulering i dag-institutionen selvstændigt forklarer, hvor godt et barn klarer sig det første år i skolen (også når der er taget højde for alle andre forhold).

Kognition, kognitive evner: Kognition betyder samlet set erkendelse og perception (opfattelses-evne). Kognitive evner kaldes i mange sammenhænge for intelligens, evnen til at løse logiske problemstillinger, men dækker mere end genetisk intelligens. Kognitive evner udvikles under graviditeten og under den tidlige og senere barndom.

Parametrisk/ikke parametrisk: I statistiske analyser er det ofte en parameter, man er interesseret i at finde frem til, at fx uddannelse forklarer indkomst. Man kan have strenge antagelser om den underliggende fordeling, der skal beskrive parameteren/parametrene. Eksempelvis at den skal være normalfordelt eller logistisk fordelt. Non-parametrisk refererer til, at man forsøger at estimere, anvende estimationsmetoder eller finde frem til forklaringer, uden så strenge antagelser om fordelinger og metoder. I stedet for at anvende strengt parametriserede estimators som den kendte maksimum likelihood estimator, kan General Method of Moments (GMM) anvendes sammen med robuste standard fejl (for at tage højde for heteroskedasticitet (varierende fejlvarians i fordelingen i data)). Denne estimationsmetode er i betydeligt omfang udviklet inden for en semiparametrisk tradition, og anvendes i stigende grad (se også Abadie, 2005).

Outcome: Defineres som udfald, eller udkomme.

Quasi-eksperimenter: Naturligt eksperiment.

Randomiseret: Tilfældig udvælgelse, stokastisk udvælgelse.

Selektionsproblem: Indebærer, at noget utilsigtet er sorteret fra, eller at et materiale er skævt.

Uobserveret heterogenitet: Ofte mangler vi oplysninger i forbindelse med gennemførelse af undersøgelser. Disse kan være korreleret, hænge sammen, med de forklarende variable eller fejlleddet. Det kan føre til, at vore estimater, koefficienter bliver biased, enten i opadgående retning eller i nedadgående ret-



ning. Nu kan vi aldrig få fuld information om en given situation, men man må statistisk set tage højde for de uobserverede forhold gennem at vælge de rigtige statistiske procedurer.

Signifikans: Statistisk betydningsfuld.

Signifikante: Statistisk betydningsfulde.

Unbiased: Koefficienter eller estimater der ikke er fejlbehæftede.

Noter

1. Tak til referee, Niels Ploug og Britha Jensen for meget nyttige og hjælpsomme kommentarer til tidligere versioner af artiklen.
2. Det er dog relevant at understrege, at der findes et bredt perspektiv på børns kompetencer (se fx Kampmann, 2003).
3. Der findes en nærmere forklaring på en række tekniske udtryk og begreber i en ordliste bagest i denne artikel. Disse udtryk er alle i teksten skrevet i kursiv som fx her kausal effekt.
4. Eksempelvis $y = f(x) = x$, hvor y er udfaldet, $f(x)$ er den stabile funktion, og x er faktoren.
5. Esping-Andersen (2004) og referencer heri, og senere McIntosh & Munk (2007b) har påpeget betydningen af daginstitutioner, der potentielt kan bidrage til at mindske den sociale reproduktion eller øge den sociale mobilitet.
6. Heckman (2000) henviser i øvrigt til et studie af Eric Hanushek, der viser, at en reduceret klassestørrelse i børnehaven faktisk har en positiv effekt på, hvor godt børnene klarer sig i en skoletest i de første skoleår, jf. det såkaldte STAR program.
7. Skelnen mellem endogene og eksogene forhold er meget kort beskrevet i ordlisten. Problemerne med at kontrollere endogene variabler er udmærket beskrevet i den noget tekniske tekst af Heckman og Navarro-Lozano (2004).
8. Mange sociologer ville mene det samme, et eksempel herpå er Pierre Bourdieus forsøg på at løse struktur-aktør problemet ved at indføre begrebet om habitus, eller systemet af dispositioner.

Litteratur

- Abadie, Alberto 2005: "Semiparametric Difference-in-Difference Estimators". *Review of Economic Studies*, 72:1-19.
- Angrist, Joshua D. 2001: "Estimation of Limited Dependent Variable Models With Dummy Endogenous Regressors: Simple Strategies for Empirical Research". *Journal of Business & Economic Statistics*, 19:2-28 (with comments from Hahn, Imbens, Moffitt, Mullhay, Todd and Reply from Angrist).
- Angrist, Joshua D. 2004: "American Education Research Changes Tack". *Oxford Review of Economic Policy*, 20 (2):198-212.
- Blundell, Richard, Dearden, Lorraine & Sianesi, Barbara 2004: *Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS*. Centre for the Economics of Education, LSE.
- Blundell, Richard, Dearden, Lorraine & Sianesi, Barbara 2005: "Evaluating the Effect of Education on Earnings: Models, Methods and Results from the National Child Development Survey". *Journal of Royal Statistical Society*, A 168, Part 3:473-512.



- Bourdieu, Pierre 1973: "Cultural Reproduction and Social Reproduction". I: Brown, R. K.: *Knowledge, Education and Cultural Change*. Papers in the Sociology of Education: 71-112. London: Tavistock.
- Bourdieu, Pierre 2000: *Pascalian Meditations*. Cambridge: Polity Press.
- Buckley, Jack & Shang, Yin 2003: "Estimating policy and program effects with observational data: the "differences-in-differences" estimator." *Practical Assessment, Research & Evaluation*, 8 (24), <http://PAREonline.net/getvn.asp?v=8&n=24>
- Campbell, Donald T. & Stanley, Julian C. 1963/1966: *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally College Publishing Company.
- Cobb-Clark, Deborah A. & Crossley, Thomas 2003: "Econometrics for Evaluations: An Introduction to Recent Developments". *The Economic Record*, 79:491-511.
- Cook, Thomas D. & Campbell, Donald T. 1979: *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Chicago, IL: Rand McNally College Publishing Company.
- Cook, Thomas D. & DeMets, David L. 2008: *Introduction to Statistical Methods for Clinical Trials*. London: Chapman & Hall/CRC.
- Cunha, Flavio, Heckman, James J., Lochner, Lance & Masterov, Dimitriy V. 2006: "Interpreting the Evidence of Life Cycle Skill Formation", I: Hanushek, E. A. & Welch, F., *Handbook of Economics of Education*: 307-458. Amsterdam: North Holland.
- Currie, Janet 2001: "Early Childhood Education Programs". *Journal of Economic Perspectives*, 15:213-238.
- Currie, Janet 2006: *The Invisible Safety Net*. Princeton: Princeton University Press.
- Currie, Janet & Thomas, Duncan 2000: "School Quality and the Long-term Effects of Head Start". *Journal of Human Resources*, XXXV: 755-774.
- Esping-Andersen, Gøsta 2004: "Untying the gordian knot of social heritage". *Research in Social Stratification and Mobility*, 21:115-138.
- Freedman, David A. 2005: *Statistical Models, Theory and Practice*. Cambridge: Cambridge University Press.
- Garces, Eliana, Currie, Janet & Thomas, Duncan 2002: "Longer-Term Effects of Head Start". *American Economic Review*, 92:999-1012.
- Greene, William H. 2003: *Econometric Analysis*. 5. udgave, Upper Saddle River, New Jersey: Prentice Hall.
- Goldberger, Arthur 1972: *Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations*. Madison, Wisconsin: University of Wisconsin Press.
- Greenberg, David H. & Morris, Stephen 2005: "Large-Scale Social Experimentation in Britain". *Evaluation*, 11:223-242.
- Heckman, James J. & Rubinstein, Yona 2001: "The Importance of Non-Cognitive Skills: Lessons from the GED Testing Program Papers and Proceedings". *American Economic Review*, 91:145-149.
- Heckman James J. & Navarro-Lozano, Salvador 2004: "Using matching, instrumental variables, and control functions to estimate economic choice models". *Review of Economics and Statistics*, 86:30-57.
- Heckman, James J. 2000: "Policies to foster human capital". *Research in Economics*, 54: 3-56.



- Heckman, James J. 2005a: "The Scientific Model of Causality". *Sociological Methodology*, 35:1-97.
- Heckman, James J. 2005b: "Rejoinder: Response to Sobel". *Sociological Methodology*, 35:135-162.
- Jespersen, Cathrine 2006: *Socialt udsatte børn i dagtilbud*. København: Socialforskningsinstituttet, rapport 06:21.
- Kampmann, Jan 2003: "Barndomssociologi – fra marginaliseret provokatør til mainstream leverandør". *Dansk Sociologi*, 14(2):79-93.
- Macdonald, Geraldine 2001: "Evaluating Effectiveness", I: Macdonald, Geraldine, *Effective Interventions for Child Abuse and Neglect. An Evidenced-based Approach to Planning and Evaluating Interventions*: 3-18, Chichester: Wiley.
- Maddala, Gangadharrao S. 1983: *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, U.K.: Cambridge University Press.
- Magnuson, Katherine A., Ruhm, Christopher J. og Waldfogel, Jane 2007: "Does Prekindergarten Improve School Preparation and Performance?" *Economics of Education Review*, 26 (1):33-51.
- Magnuson, Katherine A., Meyers, Marcia K., Ruhm, Christopher J. & Waldfogel, Jane 2004: "Inequality in Preschool Education and School Readiness". *American Educational Research Journal*, 41:115-157.
- Magnuson, Katherine A. & Waldfogel, Jane 2005: "Early Childhood Care and Education: Effects on Ethnic and Racial Gaps in School Readiness". *The Future of Children*, 15:169-196.
- Manski, Charles F. 1995: *Identification Problems in the Social Sciences*. Cambridge, MA.: Harvard University Press.
- McIntosh, James & Munk, Martin D. 2007a: "Scholastic Ability vs. Family Background in Educational Success: Evidence from Danish Sample Survey Data". *Journal of Population Economics*, 20 (1):101-120.
- McIntosh, James & Munk, Martin D. 2007b: Family Background and Educational Choices: Changes Over Five Danish Cohorts, Presented at EEA, Budapest 2007, SFI-Working Paper 18: 2007.
- Morgan, Stephen L. & Harding, David 2006: "Matching Estimators of Causal Effects, Prospects and Pitfalls in Theory and Practice". *Sociological Methods & Research*, 35:3-60.
- Morgan, Stephen L. & Winship, Christopher 2007: *Counterfactuals and Causal Inference. Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
- Rubin, Donald B. 1974: "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". *Journal of Educational Psychology*, 66:688-701.
- Shadish, William J., Cook, Thomas D. & Campbell, Donald T. 2002: *Experimental and Quasi-Experimental Designs, for generalized Causal Inference*, Houghton Mifflin College Division.
- Smith, Herbert L. 1997: "Matching with multiple controls to estimate treatment effects in observational studies". *Sociological Methodology*, 27:325-353.
- Sobel, Michael E. 1995: "Causal Inference in the Social and Behavioural Sciences", I: Armingier, G., Glogg, C.C. og Sobel, M.E., *Handbook of Statistical Modelling for the Social and Behavioural Sciences*: 1-38, New York: Plenum Press.
- Winship, Christopher & Sobel, Michael 2004: "Causal Inference in Sociological Studies", I: Hardy, M. & Bryman, A., *Handbook of Data Analysis*: 481-503, London: Sage.



Winship, Christopher & Morgan, Stephen L. 1999: "The Estimation of causal effects from observational data". *Annual Review of Sociology*, 25:659-706.

Wooldridge, Jeffrey M. 2002: *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA.: The MIT Press.

Woodward, James 2003: *Making Things Happen*. Oxford: Oxford University Press.

<http://www.alleydog.com/glossary/definition.cfm?term=Experiment>

<http://en.wikipedia.org/wiki/Experiment>

http://www.medscape.com/viewarticle/431850_2

