

**Stephen H Bell, Larry L Orr,  
John D Blomquist & Glen G Cain:  
Program Applicants as a Comparison Group in  
Evaluating Training Programs**

W E Upjohn Institute for Employment Research,  
Kalamazoo, Michigan, 1995, 181 s.

Den icke-experimentella ansatsen är den metod som använts flitigast för att utvärdera individuella effekter av arbetsmarknadsutbildning. Effekterna av utbildningen mäts med statistiska modeller och inom ramen för dessa jämförs skillnader i inkomster efter genomgången utbildning mellan deltagare och icke-deltagare. Icke-deltagare utgör jämförelsegruppen och väljs av utvärderaren med målet att de skall likna deltagare i de flesta avseenden utom deltagande i åtgärden. Detta är av naturliga skäl ofta svårt att uppnå.

Vi vet a priori att dessa två grupper kan skilja sig åt både med avseende på observerbara faktorer (t ex arbetslivserfarenhet, ålder) och icke-observerbara faktorer (t ex motivation, förmåga). Skillnader i icke-observerbara egenskaper ger upphov till selektionsproblem antingen genom att deltagare själva fattar ett beslut att delta (självselektion) eller genom att de utses av den administratör som hanterar urvalet av deltagarna (åtgärdsselektion). Besluten att delta fattas på grunder som ofta är svåra för utvärderaren att observera, men beroende på vilka antaganden som utvärderaren gör om dessa beslut går det att ta hänsyn till selektion i den statistiska modellen. Ett problem är dock att olika antaganden om selektionen innebär oli-

ka statistiska modeller, och dessa kan ibland producera diametralt motsatta slutsatser om effekten av utbildning på deltagarens framtida inkomstutveckling. Slutsatserna kan också påverkas (ibland kraftigt) av vilka icke-deltagare som deltagarna jämförs med.

Dessa två problem har medfört att många forskare numera anser att endast renodlade experiment kan fånga den "sanna" effekten av en åtgärd. I denna alternativa ansats fördelas personer slumpmässigt till utbildning eller till en kontrollgrupp som inte får utbildning. Personerna väljs från en och samma grupp och i denna ingår generellt sett enbart personer som administratörerna har bedömt som lämpliga för åtgärden. Detta förfarande innebär att grupperna av deltagare och icke-deltagare kan anses vara identiska med avseende på icke-observerbara faktorer. Det antas att den "sanna" effekten av åtgärden erhålls genom att jämföra skillnader i inkomster efter utbildning mellan grupperna. Dessa grundläggande begrepp och problem diskuteras i det inledande kapitlet i *Program Applicants as a Comparison Group in Evaluating Training Programs* skriven av Stephen H Bell, Larry L Orr, John D Blomquist och Glen G Cain. Författarna som alla har lång erfarenhet av utvärderingsforsk-

ning har skrivit en bok som på ett enkelt och tydligt sätt redogör för hur nya jämförelsegrupper kan användas till att öka precisionen i de icke-experimentella resultaten.

Det inledande kapitlet innehåller också en genomgång av erfarenheterna från 30 år av amerikansk forskning inom området. Baserat på denna genomgång drar författarna slutsatsen att det är mycket svårt att ta hänsyn till selektionsproblemen samt att tillförlitligt kontrollera för dessa även i de mest avancerade statistiska modellerna. En annan slutsats som dras är att det för närvarande varken bland forskare eller beslutsfattare finns en generell accepterad statistisk modell som kan användas för att mäta den "sanna" effekten av utbildning. Vidare argumenterar författarna att valet av jämförelsegrupp är grundproblemet i tidigare icke-experimentella studier. Jämförelsegruppen har för det mesta konstruerats som ett slumpmässigt urval från en grupp som inte är aktuell för åtgärden, eller som ett "matchat" urval där personer väljs på basen av att de har liknande observerbara karakteristika som deltagarna. Typiska kriterier för en sådan matchning har varit att icke-deltagare skall ha upplevt samma inkomstutveckling före åtgärden och vara i samma ålder eller ha samma tidigare arbetslivserfarenhet som deltagare. Författarna introducerar här ett nytt begrepp och kallar de traditionellt valda jämförelsegrupperna för "externa" jämförelsegrupper. Budskapet i det inledande kapitlet är att dessa inte är lämpliga som jämförelsenorm och i de återstående kapitlen diskuteras istället fördelarna med tre alternativa s.k. "interna" jämförelsegrupper.

Kapitel två innehåller en teoretisk motivering för att använda interna jämförelsegrupper samt en presentation av den empiriska modellen. Diskussionen är enkel och precis, och läsaren förstår snabbt skillnaden mellan externa och interna jämförelsegrupper. Den förstnämnda består av personer som antingen inte är behöriga för åtgärden eller som är behöriga men inte söker den. Det påpekas redan i det första kapitlet att de inte är lämpliga som jämförelsegrupp, men anledningen framkommer tydligare i kapitel två. Här argumenterar författarna i stället för tre interna jämförelsegrupper. Den första består av personer som är *aktuella för åtgärden* men som väljer att inte delta (*withdrawals*). Gruppen anses lämpligare eftersom utvärderaren kan förvänta sig att personer som söker men inte deltar i åtgärden liknar deltagarna i utbildning både med avseende på motivation och skäl till varför en person väljer utbildning. Den andra gruppen består av personer som har blivit *antagna till utbildningen* men som beslutar sig för att inte delta (*no-shows*). Personerna i denna grupp antas likna deltagarna i högre grad än *withdrawals*, dels på grund av att de drar sig ur senare, dels för att de har bedömts lämpliga av administratörerna. Skillnaden mellan *withdrawals* och *no-shows* är således att de förstnämnda ej värderats av administratörerna. Den tredje jämförelsegruppen består av de som har *sökt åtgärden* men som har *fått avslag* och bedömts olämpliga av administratörerna (*screen-outs*). Enligt författarna skiljer sig dessa personer från deltagare med avseende på åtgärdsrelaterade faktorer, men inte med avseende på individuella faktorer

som leder till självselektion bland *withdrawals* och *no-shows*. Man finner i detta kapitel en övertygande argumentation för användande av *screen-outs*. Gruppen anses vara lika motiverad som deltagarna i utbildning och om man i den statistiska analysen dessutom kontrollerar för de kriterier som används av administratörer för att utesluta dem från åtgärden (åtgärdsselektionen) bör man enligt författarna erhålla resultat som ligger nära de som erhålls i ett experiment. I kapitlet poängteras att utvärderaren måste ha information om hur administratörerna faktiskt väljer deltagare för att en modell baserad på *screen-outs* skall ge tillförlitliga resultat.

Kapitlet innehåller också en precisering av vad författarna anser krävs för att en icke-experimentell metod generellt sett skall kunna betraktas och accepteras som tillförlitlig. Tre krav måste ställas på en metod. För det första skall den vara operationell, dvs inte dyr att använda. För det andra skall den med rätt typ av data vara fri från en stor del av selektionsproblemen. För det tredje skall metoden vara accepterad av beslutsfattare. Den sista punkten har ofta förbisetts i den allmänna debatten och med detta konstaterande etablerar författarna tidigt en koppling mellan forskaren och beslutsfattaren. Denna koppling behåller författarna i de återstående kapitlen. Detta är intressant och ett tillskott till den allmänna debatten om metoders användbarhet.

I kapitel tre analyseras en amerikansk utbildningsåtgärd. Den data som använts innehåller information både om interna jämförelsegrupper och en renodlad experimentell kontrollgrupp som består av personer som har accep-

terats av administratörerna men slumpmässigt nekats deltagande. Den sanna effekten av åtgärden erhålls genom att jämföra inkomster mellan deltagare och kontrollgruppen. Detta resultat är centralt, för genom att jämföra det med de resultat som erhålls med de interna jämförelsegrupperna kan man granska hur effektivt dessa kontrollerar för individ- och åtgärdsselektion. Resultaten för de interna jämförelsegrupperna har uppmänts inom ramen för en enkel icke-experimentell modell. Enligt författarna är valet av modell ointressant då målet är att värdera betydelsen av olika jämförelsegrupper.

Kapitlet innehåller en mängd resultat som varierar beroende på vilken jämförelsegrupp som används och vilka utfallsår som studeras. Tittar man t ex på resultaten tre respektive fyra år efter genomförd utbildning finner man en positiv löneeffekt, samt att modeller baserade på *screen-outs* producerar resultat som ligger närmast de "sanna" resultaten. Skillnaden är ungefär \$ 300. Jämför man grupperna på samma sätt både före och efter åtgärden så ger modeller baserade på *no-shows* liknande resultat som modeller där *screen-outs* används som jämförelsenorm. Enligt författarna tyder detta på att båda grupperna utgör potentiella alternativ till de traditionella jämförelsegrupperna som oftast består av personer som inte har sökt åtgärden. Det mest entydiga resultatet i kapitel tre är att *withdrawals* inte är en tillförlitlig jämförelsegrupp oavsett vilken tidsperiod som studeras.

I kapitel 4 diskuteras ingående en metod som kan användas för att bedöma hur stort selektionsproblemet är och som kan avgöra vilka resultat som är

rimligast. Metoden jämför olika statistiska fördelningar, och trots att det är ett tekniskt komplicerat problem inom utvärderingsforskningen är kapitlet mycket lättillgängligt. Givet de statistiska fördelningarna diskuteras också vilken typ av beslut som kan fattas, samt inom vilka gränser som man kan fatta ett korrekt beslut. Författarna relaterar resultaten till olika hypotetiska problem som beslutsfattarna kan stå inför. Vidare diskuteras hur stora selektionsproblem som beslutsfattare kan acceptera för att en skattad effekt skall betraktas som trovärdig. Författarna visar tydligt att de underliggande administrativa beslutsreglerna har stor betydelse för de faktiska slutsatser som kan dras från icke-experimentella resultat. De visar också att man inom ramen för olika regler kan tillåta en viss grad av selektionsproblem utan att man för den skull behöver dra felaktiga slutsatser. Detta mynnar ut i konstaterandet, att om man vet vad beslutsfattarna tillåter och vilka beslutsregler de följer, är det möjligt att testa om de icke-experimentella skattningarna uppfyller just de kraven. Detta förutsätter att man känner till den sanna effekten av åtgärden för endast då går det att fastställa graden av selektionsproblem i de uppmätta inkomsteffekterna.

Resultaten från testen som jämför de icke-experimentella resultaten med det "sanna" resultatet leder tyvärr till slutsatsen att ingen jämförelsegrupp kan betraktas som ett lämpligt substitut för en experimentell kontrollgrupp. Författarna finner t ex att det rimligaste resultatet avviker med mer än femtio procent från det "sanna" experimentella resultatet. Trots det argumenterar för-

fattarna för att *screen-outs* och *no-shows* är lämpligare än de jämförelsegrupper som normalt används. De anknyter här till den teoretiska diskussionen i kapitel två. Författarna avslutar kapitlet med att föreslå fler test av ansatsen. Vidare rekommenderas utvärderaren att samla in mycket information både om faktorer som styr den administrativa selektionen av deltagare och om faktorer som gör att accepterade deltagare inte deltar i åtgärden. Sådan information anses råda bot på mycket av de grundläggande selektionsproblemen.

Boken sammanfattas i kapitel fem. Kapitlet ger också vissa rekommendationer och bl a föreslår författarna att så länge man ej kan genomföra experimentella utvärderingar bör utvärderaren använda någon av de två föreslagna jämförelsegrupperna, och då helst "screen-outs". Detta kan man finna bra teoretiskt men inte helt övertygande empiriskt stöd för. Därför anser jag att rekommendationen bör tolkas i ett vidare perspektiv, som att det är viktigt att analysera känsligheten i resultaten för alternativa jämförelsegrupper. Detta utesluter varken de föreslagna "interna" eller de förkastade "externa" jämförelsegrupperna.

Som helhet betraktat innehåller boken en bra sammanfattning av de problem som föreligger vid användandet av statistiska modeller för utvärdering av arbetsmarknadspolitiska åtgärder. Författarna redogör och argumenterar också övertygande för några alternativ till de jämförelsegrupper som vanligen utnyttjas. Just denna övertygande diskussion innebär att betydelsen av studier som inte riktigt överensstämmer med den bild författarna vill förmedla tonas

ned. Till exempel har James Heckman och Joseph Hotz (1989) visat att det är möjligt att hitta det icke-experimentella resultat bland många som ligger närmast det som har erhållits med en rent experimentell ansats. Studien använder dessutom "externa" jämförelsegrupper. Vidare visar en annan studie av James Heckman och Rebecka Roselieus (1994), att om man skapar en jämförelsegrupp från en grupp av personer som är behöriga att delta och som dessutom befinner sig på samma lokala arbetsmarknader som deltagare, kan man få resultat som ligger nära de som erhålls med ett slumpmässigt experiment. Resultatet tyder således på att "matchade" jämförelsegrupper i vissa fall kan utgöra en lämplig jämförelsenorm. Båda referenserna finns i bokens litteraturförteckning.

En annan kritik som man kan rikta mot boken är att det inte redogörs för de underliggande antaganden som måste vara uppfyllda för att de interna jämförelsegrupperna verkligen skall utgöra en lämplig jämförelsenorm. Till exempel nämns inte antagandet att jämförelsegrupperna inte skall ha erhållit ett närliggande substitut till den åtgärd som utvärderas, vilket är standard i litteraturen.

I boken betonas betydelsen av administratörernas selektionskriterier men författarna diskuterar inte svårigheterna med att identifiera hur administratörerna väljer deltagare. Vidare innehåller boken inte en diskussion om hur utvärderaren skall kunna konstruera lämpliga mått på åtgärdsselektionen. En sådan diskussion är naturlig dels för att måttet framställs som centralt, dels för att det är viktig information för utvärderare som vill använda metoden.

Den allmänna utgångspunkten i boken är att ett slumpmässigt experiment producerar ett "sant" mått på utbildningens effekt på framtida inkomster. Trots det diskuteras inte nya studier som har visat att också experiment kan ge missvisande resultat.

Denna kritik gör inte boken mindre intressant utan visar enbart att författarna argumenterar starkt för en alternativ ansats som kan vara fruktbar på lång sikt. Dessutom innehåller boken en bra beskrivning av problem som utvärderare ofta stöter på.

**Håkan Regné** är doktorand i nationalekonomi vid Institutet för social forskning Stockholms universitet