

Per Johansson och Sophie Langenskiöld

Replik till Åke Dahlerg

I förra numret av Arbetsmarknad & Arbetsliv framförde Åke Dahlerg kritik mot vår utvärdering av ”Arbetsstorget för erfarna” (Johansson & Langenskiöld 2008). Han avslutar sin artikel med att förespråka ”experiment eller en kombination mellan en ekonometrisk och en beteendevetenskaplig ansats” (Dahlerg 2008 s 78). På detta allmänna plan har vi naturligtvis inga invändningar. Även om randomiserade experiment har sina nackdelar, så propagerar IFAU (Institutet för arbetsmarknadspolitisk utvärdering) oftast för experiment som utvärderingsmetod. Det finns heller ingen motsättning mellan en effektstudie och användandet av kvalitativa data. I själva verket berikar dessa ansatser varandra.

Men i andra avseenden framstår Åke Dahlergs kritik mot vår studie som grundlös och inkonsekvent. Innan vi bemöter Åke Dahlergs sakkritik finns det dock anledning att reda ut begreppen lite grann. Vi börjar därför med att diskutera värdet av experiment i en utvärderingssituation.

Om experiment som utvärderingsmetod

Den stora fördelen med ett experiment är att vi vet varför en individ har fått del av en åtgärd. Detta medför att experiment – i idealfallet – ger resultat som är internt valida, det vill säga de visar på effekten för de deltagare som deltog i åtgärden (till exempel äldre långtidsarbetslösa i Stockholm). Att resultaten har intern validitet innebär dock inte att de med nödvändighet har extern validitet. Resultat som inte har extern validitet kan inte överföras till andra populationer (till exempel äldre långtidsarbetslösa i Sveg) än den som faktiskt deltog i åtgärden.

Detta är en idealbild av vad experiment kan åstadkomma. I praktiska tillämpningar kan randomiserade studier vara förknippade med problem (se till exempel Heckman

med flera 1999). Ibland kan både kontroll- och behandlingsgrupp vara kontaminerad; till exempel kan individer som ingått i kontrollgruppen fått del av en annan alternativ behandling. I vissa fall uppkommer även problem som har att

Per Johansson är professor i ekonometri, särskilt mikroekonomisk utvärderingsforskning vid nationalekonomi, Uppsala universitet och verksam vid Institutet för Arbetsmarknadspolitisk Utvärdering (IFAU).
per.johansson@ifau.uu.se

Sophie Langenskiöld är doktor i ekonomi och adjungerad forskare vid Institutet för Arbetsmarknadspolitisk Utvärdering (IFAU).
sophie.langenskiöld@ifau.uu.se

göra med att kontroll- och behandlingsgrupp har behandlats på olika sätt, även i andra avseenden än dem som har att göra med själva åtgärden.

Dessutom är experiment inte alltid genomförbara, både av etiska skäl eller kostnadsskäl. Även om det är en god regel att verka för att genomföra randomiserade studier, tvingas utvärderare allt som oftast fundera på vilka alternativ som ger trovärdiga resultat.

Alternativ till experiment

Av flera skäl är det viktigt att fundera på hur en trovärdig utvärdering av ett program ska kunna genomföras *innan* programmet sätts i stället för efter (som i fallet med Arbetstorget för erfarna, AE).

Det väsentliga är att beslutsfattare tänker igenom hur en åtgärd ska utvärderas innan man genomför åtgärden. Man kan tänka sig en mängd olika utvärderingsstrategier och vi ger två exempel nedan.

1. Med ett klart och tydligt regelsystem som avgör vem som får åtgärden och vem som inte får åtgärden blir självselektion eller selektion av handläggare ett mindre problem och åtgärden kan utvärderas på ett trovärdigt sätt med hjälp av statistiska metoder. Nedan ges tre konkreta exempel på design.
 - a. Sekventiellt introducera den nya åtgärden för olika kontor.
 - b. Ge åtgärden deterministiskt till alla långtidsarbetslösa 55 år eller äldre men inte till några yngre än 55 år.
 - c. Ge ett erbjudande om åtgärden till arbetslösa äldre än 55 år slumpmässigt vid olika arbetslöshetstider så länge de fortfarande är inskrivna som arbetslösa vid en arbetsförmedling.
2. Genomföra prospektiva studier. Detta innebär att man samlar in detaljerade uppgifter (till exempel via enkäter) om arbetslösa personers motivation, sökbeende, hälsa et cetera innan införandet av åtgärden i fråga. Sedan samlar man in uppgifter vid ett senare tillfälle. Effektutvärderingen består då i att mäta förändringarna i dessa självrapporterade uppgifter för dem som gick in i åtgärden och ställa dem mot förändringarna i gruppen som inte deltog i åtgärden.

Men i praktiken är det få beslutsfattare som tänker på utvärderingen innan de fattar beslut om att införa en åtgärd. I dessa fall får man förlita sig på retrospektivt insamlade data. Detta gör det svårare att utvärdera programåtgärder, vare sig informationen är insamlad via enkäter, intervjuer eller registerdata.

Matchning som alternativ till experiment

I de allra flesta utvärderingssituationer ställs utvärderaren inför faktum och har således bara retrospektiva data till sitt förfogande. En rimlig ansats är då att för-

söka efterlikna en experimentsituation så långt som möjligt. Matchning är ett sätt att skapa en sådan situation. Ambitionen i en matchningsstudie är normalt att skatta en effekt som gäller för dem som faktiskt deltagit i åtgärden.

Matchning är baserat på följande grundantagande: givet en uppsättning observerade egenskaper (till exempel arbetslöshetshistorik, utbildning och ålder), antas åtgärdsdeltagandet vara oberoende av de icke-observerade faktorer som kan påverka framgången på arbetsmarknaden (det är som om slumpen har avgjort vem som har fått del av åtgärden). Under detta antagande kan vi skatta effekten av åtgärden genom att jämföra två individer – en deltagande individ och en jämförelseindivid – som båda har varit arbetslösa lika länge, samt har samma utbildning och ålder. Fördelen med matchningsansatsen jämfört med till exempel regressionsanalys är att man skapar en kontrollgrupp utan att studera utfallsvariabeln (till exempel sökintensitet, hälsa, arbete et cetera) och detta förhindrar ”sökande” efter effekter. När man väl bestämt vad som är en lämplig kontrollgrupp (det vill säga motsvarande det som det randomiserade experimentet skulle skapa) skattas effekten som skillnaden mellan de behandlades medelvärden och kontrollgruppens medelvärden.

Trovärdigheten i all typ av utvärdering handlar i slutändan om hur rimlig jämförelsen som ligger till grund för effektskattningen är. Trovärdigheten i en matchningsstudie (i likhet med all empirisk forskning) beror till syvende och sist på hur mycket man kan observera om individerna.

Matchning som metod i utvärderingen av Arbetstorget för erfarna

I denna utvärdering kontaktades vi av Åke Dahlberg efter det att studien hade påbörjats för att genomföra en effektstudie. Med tanke på det specifika selektionsförfarandet av deltagare till programmet var det inte möjligt att genomföra en trovärdig effektstudie bara utifrån retrospektivt insamlade registeruppgifter. Vi beslutade därför gemensamt att genomföra en enkätundersökning där vi samlade in uppgifter om till exempel sökbeteende, motivation, tidigare erfarenhet och hälsa både före och efter programmet. Ett problem är att uppgifterna som avser tiden innan åtgärdsstart mäts i efterhand (retrospektivt) vilket gör att svaren blir mindre tillförlitliga än om de mäts innan (prospektivt) åtgärden börjat (jämför punkt 2 ovan).

Arbetstorget för erfarna beslutade att bekosta enkätundersökningen. I samarbete med Arbetstorget för erfarna utformades enkäten och testades vid tre olika fokusgruppsmöten. Den ansågs av alla berörda parter (företrädare för Arbetstorget för erfarna, Åke Dahlberg och av oss) fungera för att genomföra en trovärdig utvärdering av programmet.

Ambitionen i detta fall var att skatta effekten av programmet för deltagarna. Eftersom programmet var till för äldre långtidsarbetslösa i Stockholm är det na-

turligt att välja ut jämförelseindivider ur samma population. Enkäten skickades ut till deltagarna i Arbetstorget för erfarna och jämförelseindivider som bodde i Stockholms kommun hade samma fördelning på kön, ålder och arbetslöshetshistorik, och var arbetslösa vid samma tidpunkt som programdeltagarna. Eftersom vi trodde att svarsfrekvensen skulle vara lägre i jämförelsegruppen valde vi att ta ett större urval av arbetslösa icke-deltagare. För varje deltagare valde vi fem icke-deltagare. Per konstruktion kommer inte bortfallet att vara skevt i termer av observerade egenskaper och arbetslöshetshistorik. I tillägg konstanthåller vi för deltagarnas och jämförelseindividuernas självrapporterade motivation, hälsa och sökbeteende innan programmet startade. Det grundläggande antagandet för utvärderingen är alltså att individer med samma observerade egenskaper, arbetslöshetshistorik, motivation, sökbeteende et cetera är jämförbara. Bortfallet är inget problem för den interna validiteten under förutsättning att vi kan tro på de grundläggande antagandena.

Mer om Åke Dahlbergs sakkritik

Enligt egen utsaga är Åke Dahlberg inte kritisk till vår rapport utan kritisk till den typ av kvantitativ utvärderingsstudie som vår rapport representerar. Han vill därför ha en diskussion om den metod som använts. Vi tror dock inte riktigt på att Åke Dahlberg på allvar är intresserad av en sådan diskussion utan drivs av andra incitament än rent vetenskapliga. Vi finner i synnerhet fem sakförhållanden problematiska.

1. Åke Dahlberg och IFAU hade intentionen att skriva en gemensam studie ända fram till dess att resultaten blev kända. När resultaten blev kända bestämdes att Åke Dahlberg skulle skriva en separat rapport. Denna rapport består till stora delar av kritik mot vår rapport. Vi välkomnar konstruktiv kritik, men förfarandet framstår som anmärkningsvärt, särskilt när alla parter var överens om utvärderingsstrategin på förhand.
2. Det är märkligt att i ena stunden diskutera det svåra selektionsproblemet (Dahlberg 2008 s 77) för att sedan helt bortse från denna problematik. Det är denna typ av logiska hopp som Åke Dahlberg gör sig skyldig till när han använder sig av aggregerad statistik (tabell 3 och 4) för långtidsarbetslösa personer i ålder 50-64 i hela landet för att påvisa att AE har en effekt. Åke Dahlberg behöver förklara varför denna jämförelse är rimlig. Varför skulle äldre arbetslösa individer i till exempel Sveg vara jämförbara med programdeltagare i Stockholm? Den jämförelse som Åke Dahlberg gör är orimlig.
3. Det är anmärkningsvärt att så mycket energi läggs på att argumentera för att sysselsättningsresultatet i vår rapport inte är trovärdigt. Däremot

verkar Åke Dahlberg inte ha några problem med övriga resultat – ökad sökintensitet, ökad kontaktfrekvens och förbättrad (självupplevd) hälsa – som är framtagna med samma metod.

4. Att i ena stunden förespråka experiment (som i grunden är en helt teorilös ansats) för att i andra stunden använda sig av sökteori för att förkasta empiriska resultat är inkonsekvent. Vi fann det också förvånande att inte den ökade sökintensiteten ökar sysselsättningen. Men från detta följer inte att man ska misstro sysselsättningsresultatet. Effekten på sökintensitet kanske inte är tillräcklig för att den ska få avtryck i en ökad sysselsättning. Man kan också argumentera för att sysselsättningsresultatet ska ges särskild stor tilltro eftersom vi får samma resultat på sysselsättningen från två oberoende datakällor: via register (Ams arbetslöshetsregister) och enkäten. Dessutom har retrospektiva frågor välkända problem som vi delvis redan har varit inne på. Det är inte långsökt att tänka sig att individer som uppenbarligen är nöjda med programmet (enligt Åke Dahlbergs intervjuundersökning) i efterhand också överskattar dess effekt på till exempel sökintensiteten.
5. Det är klart att intervjuer är ett mycket värdefullt komplement till kvantitativa effektstudier. De är dock ingalunda ett substitut. Intervjuer med ett 30-tal deltagare och ett 15-tal projektmedarbetare är en svag grund för att uttala sig generellt om åtgärdens effekter och tillika en svag grund för att förkasta effektskattningar som inte passar i Åke Dahlbergs föreställningsvärld.

Sammanfattningsvis har vi alltså mycket svårt att se att vissa delmängder av resultaten ska ges en större tilltro än andra. Vår inställning är att antingen får man tillstå att alla resultat är trovärdiga eller så får man argumentera för att inga är det. Den senare ståndpunkten är respektabel men innebär att vi bara kan förlita oss på experimentell evidens (om ens detta). I alla sammanhang skulle vi hävda att studier som försöker kontrollera för observerade egenskaper är avsevärt trovärdigare än studier som grundar sina slutsatser på skillnader i medelvärden.

Referenser

- Dahlberg Å (2008): "Ger inte intensifierade arbetsförmedlingsåtgärder för äldre några sysselsättnings effekter?" *Arbetsmarknad & Arbetsliv*, vol 14, nr 3, s 67-78.
- Heckman J, LaLonde R & Smith J (1999): "The Economics and Econometrics of Active Labor Market Programs." I Ashenfelter O & Card D (red): *Handbook of Labor Economics*, vol 3A (s 1865-2097). North-Holland: Amsterdam.
- Johansson P & Langenskiöld S (2008): Ett alternativt program för äldre långtidsarbetslösa – utvärdering av Arbetstorget för erfarna. Rapport 2008:2. Uppsala: IFAU.

