

Skattning av insatsers effekter i platsbaserade randomiserade försök

Robert Boruch, Ellen Foley, och Jeremy Grimshaw

1. Introduktion

Ett platsbaserat försök avser här en studie i vilken ett antal platser eller organisationer är slumpmässigt uttagna till en av två eller flera interventioner för att utröna vilken intervention som fungerar bäst. "Platserna" kan vara byar eller grannskap, skolor eller ungdomsgårdar, bostadsområden eller andra organisationer. Områden som är utvalda för interventioner är inte omedelbart urskiljbara. Statistiskt är de likvärdiga med hänsyn till det slumpmässiga urvalet. Denna likvärdighet tillåter en rimlig jämförelse, dvs en förutsättningslös bedömning av interventioners relativa effekter och ett statistiskt uttryck av forskarens tilltro till resultaten.

Försök där individer slumpmässigt blir föremål för olika interventioner är vanliga i medicinsk och annan forskning. Slumpmässig uttagning av enheter som platser och institutioner är mindre vanliga. Som Donald T. Campbell hävdade i "Reforms as Experiments":

Där tillvägagångssättet administreras genom individuell klientkontakt, kan slumpmässig uttagning på personnivå ofta åstadkommas obemärkt....

Men för de flesta sociala reformer blir större administrativa enheter inblandade, såsom klassrum, skolor, län eller delstater. Vi behöver utveckla politiska ståndpunkter och ideologier som möjliggör slumpmässig uttagning på den här nivån. (Campbell, 1969; Campbell, 1988)

Campbell har inte djupgående betraktat användning av platser eller enheter i randomiserade försök, därför att sådana försök var ovanliga på den tiden. I det som följer är vi hänvisade till Campbells insikt och bygger på andras senare arbeten. Teman är nära förknippade med utvärdering av komplexa sociala program som är avsedda att främja hälsa och välmående, välfärd och utbildning, och att minska brottslighet.

1.1 Definitioner

En uttagen enhet ("unit of allocation") syftar på vem eller vad som är slumpmässigt uttaget för olika interventioner i ett försök. Konventionella läroböcker i psykologi och utformningar av medicinska försök, behandlar experiment där individer är uttagna enheter. Här fokuserar vi på platser, administrativa enheter eller grupper istället för individer. Vi hänvisar till vad som står om platsbaserade randomiserade försök i denna artikel. Dessa kallas också för grupp-slumpmässiga försök (place-based randomized trials) (Murray, 1998) och "samling slumpmässiga försök" ("cluster randomized trials") (Donner och Klar, 2000).

Analysenheter är sådana för vilka data är tillgängliga och användbara.

Ungdomsinstitutioner kan vara slumpmässigt uttagna enheter i ett försök där man jämför två institutioners tillvägagångssätt för att minska återfall i brottslighet. Analysenheter kan vara institutioner eller både institutioner och individer inom institutioner.

1.2 Denna artikels innehåll

I de följande avsnitten diskuterar vi förutsättningar för användningen av randomiserade försök och dess berättigande samt vissa svårigheter i deras användning. Exempelen i denna uppsats är mångfaldiga, delvis för att visa att lämpliga försök kan göras i skiftande miljöer.

1.3 Förutsättningar

Den första förutsättningen är att offentliga myndigheter och privata stiftelser är intresserade av att bedöma den relativa effekten av ett nytt program som de finansierar. Uttryckt på ett annat sätt, vi förutsätter att allmänheten är intresserad av svaret på frågan: "Vad fungerar bättre, för vem och för hur länge?"

En annan förutsättning är att en försvarbar bedömning av förnyelsens effekter beror på fastställande av hur enheter skulle ha fungerat utan förnyelse. I praktiken kan man göra en sådan

bedömning utifrån t ex tidsseriestudier. Kuusis (1957) studie över effekter av alkoholförsäljning i Finland är ett anmärkningsvärt prejudikat i användningen av administrativa register i korttidsserier. Här förutsätter vi att data från tidsserier och ad hoc-jämförelser är otillräckliga för objektiv bedömning av programmets effekt. Vissa alternativ till slumpmässiga försök, däribland tidsserier och deras sårbarhet, är behandlade av Campbell och Stanley (1963) och Shadish, Cook och Campbell (2002).

Viktigast vid sammansättningen av en jämförelsegrupp är en enkel och vetenskapligt försvarbar metod, en metod som möjliggör en rimlig bedömning av relativa skillnader mellan olika program, dvs. randomisering. Till exempel kan en grupp ungdomsinstitutioner slumpmässigt väljas ut från en samling lämpliga institutioner och engageras i ett nytt interventionsprogram. Resultat från de här institutionerna skulle sedan jämföras med institutioner som slumpmässigt uttagits till fortsatt medverkan i pågående program. Slumpmässig uttagning säkerställer att de här två grupperna inte skiljer sig åt systematiskt, bortsett från inverkan av interventionsprogrammet i undersökningen.

En tredje förutsättning är att framtiden för effektutvärdering i många länder finns i kontrollerade försök som utförs i en liten skala för att undersöka vilka program som fungerar innan de genomförs på nationell eller regional nivå. Sådana försök genomförs i ökande omfattning. Boruch och Foley (2000), bland andra, nämner över 50 olika undersökningar som omfattar samhällen eller geografiska områden, skolor eller klassrum, bostadsområde och andra organisationer som uttagna enheter i en randomiserad fältstudie. (Se Boruch, 1997; Donner, 2000 och Murray, 1998 generellt, och Campbell Collaboration's Social, Psychological, Educational, and Criminological Trials Register; www.campbellcollaboration.org.)

2. Grundval: Varför använda platser som enheter?

Varför skulle vi betrakta platser eller andra enheter som randomiseringsenheter i utvärdering av programeffekter? Några anledningar är programteori, lag och etik, policy, rådgivandegruppers anvisningar, statistisk teori och bevisföringsregler.

2.1 Programteori

Med "teori" menar vi här hur en intervention förväntas att ha de effekter som vi tror den skulle ha. Uttryckt på ett annat sätt, teoretikern föreslår en "logisk modell" för att försöksvis förklara vad som händer när ett program är genomfört. Eller, forskaren kan skissera en formell modell eller en orsakskedja.

Många teorier om samhällsliga förändringar antar att ett program skulle fungera om det genomförs på ett sätt där alla delar av organisationen agerar i samförstånd. T ex forskning i förebyggande av sexuellt överförda sjukdomar beror på teorier om vilka institutionella faktorer och grupp faktorer som påverkar riskbeteende. Se Wasserhit, Aral, Holmes och Hitchcock (1991) allmänt och Hornik (1991) speciellt. Slumpmässiga fältförsök i Kalifornien och Texas har använt 20 skolor som uttagna enheter och analysenheter för att testa program som är baserade på många sådana teorier (Coyle et al., 1996; Basen-Engquist et al., 1997).

En mängd platsbaserade slumpmässiga försök har också använt skolor för att bedöma teoridrivna program som var avsedda att förebygga eller minska drogmissbruk. The Midwestern Prevention Project (Pentz, 1994) t ex, var baserat på en teori om att tonåringars droganvändning beror på faktorer som tidigare droganvändning och tonåringens skicklighet att hantera jämnårigas påtryckningar. I denna hävdas också att faktorer utanför individen som miljö och situation är viktiga därför att till exempel samhällsnormer kan påverka beteendet i tonåren.

Programteorin har också bidragit till flerstegsforskning i hur man kan engagera och uppmuntra mentalsjukhus i metoder som har visat sig vara mer effektiva i att bota vissa former av mental sjukdom. En sådan teori innefattade idéer om på vilken nivå medarbetare på ett sjukhus kan bli engagerade från början (uppifrån och ner eller nerifrån och upp) och det bästa angreppssättet. Det senare innefattade deltagande i seminarier istället för att bara skicka ut broschyrer. Förhoppningen var att folk skulle reagera olika på de här skilda angreppssätten. (Fairweather et al., 1974; Fairweather och Tornatzky, 1977).

2.2 Lag, etik och kultur

En anledning till varför platser kan användas som slumpmässigt uttagna enheter är att slumpmässig uttagning av individer i försöksprogram *inom* en plats är olaglig och oetisk. Eller att den här typen av randomisering kan vara kulturellt eller politiskt oacceptabel. Randomisering av hela platser till alternativa program kan betraktas som juridiskt och etiskt vederhäftigt.

Till exempel, i ett slumpmässigt försök som testade en modell för drogmotstånd (The Drug Abuse Resistance Education Model, D.A.R.E.) valde forskare slumpmässigt ut hela skolor som behandlings- och kontrollgrupper delvis på grund av att det hade varit svårare att få skolorna att samarbeta om endast några av deras elever erbjudits programmet (Curtin, personlig kontakt, April 3, 1996). En form av institutionell etik och kultur rådde. Att använda skolorna som slumpmässigt uttagna enheter hjälpte till att säkra kontrollskolornas medverkan i undersökningen. Skolorna i kontrollgruppen lovades tillgång till D.A.R.E.-program i ett år efter studiens avslutning.

Likaledes skulle ungdomsinstitutionerna i Sverige, till exempel, kunna ha invändningar mot randomisering av deras klienter till olika program för att kunna bestämma vilket program som är effektivast för att minska återfall. Andra etiska värderingar i den lokala enheten kan ha

företrädare, t ex att ge "samma" hjälp till alla i enheten. Ett randomiserat försök i vilket alla lämpliga och villiga enheter provar en eller två olika metoder kan betraktas som mer rimlig. Detta poängterades av Karin Tengvald i ett möte i Stockholm om utvärdering av social tjänstens program (Soydan, 1998).

Betoningen här ligger på jämförelsen mellan alternativa interventioner i olika samhällen, och inte att ge "behandling" till en grupp och försumma en annan grupp. Huvudsaken här är inte bara om behandlingen fungerar men vilken behandling som fungerar bäst.

2.3 Policy och Politik

Som policy och av politiska skäl, har offentliga myndigheter eller stiftelser som finansierar program satt bestämmelser som direkt påverkar organisationer snarare än individer. Sådana regler kräver att platser eller organisationer vidtar vissa åtgärder, ordnar förhandlingar och så vidare. Det innebär att en undersökning av programeffekter måste betrakta platsen som främsta mål i utvärderingssyfte. Individer inom platser är de yttersta målen.

Till exempel, den amerikanska regeringens policy i demonstrationsprojekt i USA har många gånger betonat att lokala samhällen är mycket viktiga när det gäller att minska sociala problem. Förebyggande av drogmissbruk är ett bra exempel. The Center for Substance Abuse Prevention (CSAP) inrättades för att minska förekomsten av alkohol, tobak och droganvändning. Centret har försökt att åstadkomma detta genom satsningar som Community Partnership Demonstration Program som har riktat in sig på att lära hur olika samhällsbaserade organisationer kan engageras i effektiv intervention. Olika sätt att göra detta förklarades av Kaftarian och Hansen (1994). Betoningen låg på lokala samhällen som randomiserade analysenheter i fältförsök (Pentz, 1994, Wagenaar et al., Ellickson, 1994, Murray och Wolfinger, 1994, Lorion, 1994).

Andra exempel på program där den mest direkta anknytningen finns mellan enheter och staten eller stiftelser, och *inte* mellan individer och bistånd, är lätta att identifiera. De framträder i kompensationsutbildning och andra program finansierade av amerikanska utbildningsdepartementet och amerikanska hälsovårds- och socialdepartementet. Lån från Världsbanken till regeringar verkar via organisationer som banker, lantbruksstationer eller skolor. Världsbanken stödjer sällan randomiserade försök, men det finns några exempel på program som är finansierade genom banklån och som har gjort undersökningar i platsbaserat försök.

2.4 Statistisk Teori och Analys

Moderna statistiska analysmetoder bygger på antaganden att en observation på vilken individ som helst eller en enhet är oberoende av observationer av alla andra. När antagandet inte håller, och analytikern undgår att inse detta, kommer analysen att bli äventyrad. Till exempel kan skillnader i programresultat förklaras statistiskt signifikant som en följd av felaktig bedömning av ickeberoende. Se till exempel Donner och Klar (2000) och Murray (1998).

Att anta att observationsenheter är oberoende är i många situationer inte rimligt. Till exempel, en viss gängmedlems respons på ett brottsförebyggande program behöver inte vara oberoende av andra gängmedlemmars respons även om programmet enbart omfattar vissa medlemmar. Ett barns poäng i ett prov som testar möjligheten att arbeta i lag är troligen inte oberoende av poäng som andra barn i samma lag fick.

För statistikern innebär allt detta att det inte är individer som slumpmässigt skall uttagas till program. Och det är inte data på individnivå som vanligtvis skall användas för att bedöma programmets effekt. Snarare skall randomisering och analys först rikta in sig på hela grupper eller organisationer och sedan på individer inom varje grupp eller enhet.

2.5 Rådgivande Grupper roll i Forsknings- och Utvärderingspolicy

I förebyggandet av farliga sjukdomar, däribland sexuellt överförbara, krävs ibland att programmen genomförs via organisationer eller geopolitiska jurisdiktioner. För detta ändamål föreslog National Academy of Science Panel on Evaluating AIDS Prevention Programs att diagnostik- och rådgivningsställen blir betraktade som enheter i kontrollerade experiment för att uppnå en förbättrad service (Coyle, Boruch och Turner, 1991). Flervetenskapliga konferenser om sexuellt överförda sjukdomar, finansierade av National Institute on Allergy and Infectious Diseases (NIAD), har lett till insikten om att kliniker, fabriker, kyrkor och andra organisationer tillsammans med lokala samhällen lämpligen kan tjäna som enheter i randomiserade försök (Green och Washington, 1991).

I övervägande av metoder för att förebygga läkemedelsmissbruk, uttalade deltagarna i Communities that Care under sin utvärderings- och designkonferens följande:

Noggrann utvärdering av en omfattande samhällsintervention kräver en undersökningsmodell där samhällen är slumpmässigt uttagna till experiment- och kontrollgrupper.

(Se Peterson, Hawkins och Catalano, 1991). I England har Joseph Rowntree Foundation påverkats av motsvarande tankar (Farrington, 1997).

National Research Council's Panel on the Understanding and Control of Violent Behavior formulerade följande rekommendation:

Panelen anmodar att uppmana till ett nytt multi-lokalsamhällesprogram för utvecklingsstudier av aggressivt, våldsamt och asocialt beteende, avsett för bättre orsaksmässig förståelse och förebyggande interventioner... (s. 25)

Reiss och Roth (1993), hävdade som författare till den här panelrapporten att "Randomiserade kontrollerade fältundersökningar ofta har viktiga fördelar som utvärderingsstrategi" (s. 320).

Slutligen, tänk på att "Design and Analysis Issues in Community Trials" var huvudtema i 1992 års National Institutes of Health-konferens. Deltagarna var eniga om att användningen av samhällen som uttagnings- och analysenheter ställde utmaningar, men att det fanns olika tekniker som kunde övervinna de här utmaningarna (Murray et al., 1994).

3. Några Exempel

Människor inser ofta inte att det är möjligt att genomföra randomiserade försök som använder organisationer eller andra enheter som tillåter rimlig jämförelse. I följande exempel, vill vi ge bevis på genomförbarhet av randomiserade försök.

3.1 Skolor, skoldistrikt och klassrum som slumpmässigt uttagna enheter

Bl a har skolor och klassrum blivit slumpmässigt uttagna till olika metoder för undervisning av barn för förebyggande av substansmissbruk (Schaps et al., 1982; Moskovits, 1984; Botvin et al, 1995; Murray, Moskovits och Dent, 1996). I försök som Drug Abuse Resistance Education, D.A.R.E. har man i Illinois randomiserat deltagare från 12 matchade skolor till olika program för att främja rimlig jämförelse (Rosenbaum et al., 1991). Flay et al.s (1985) arbete i Kanada är ett utomordentlig precedensfall på den här arenan.

I ansträngningarna att utvärdera ett teoridrivet program för reducering av alkoholanvändning bland minderåriga ungdomar, genomförde Wagenaar et al. (1994) ett randomiserat fältförsök som omfattade 15 skoldistrikt i Minnesota och Wisconsin. Sju av dem var slumpmässigt uttagna att använda ett särskilt samhällsbaserat förebyggande program. De övriga var slumpmässigt uttagna som kontrollgrupp.

Skolor har också varit utvalda enheter i minst två röningsförebyggandeförsök. The Television, School and Family Smoking Prevention Project, använde multi-attributbalansering

för att slumpmässigt utse 35 skolor i Los Angeles- området till olika mediabaserade rönkningsförebyggande kampanjer. Flay et al. (1985) utsåg slumpmässigt 22 matchade skolor till försöks- och kontrollfunktioner i Waterloo Study, en kanadensisk rönkningsförebyggande satsning. Försök som omfattade hela skolan i ett projekt för att minska risken för hjärt-kärlsjukdomar har också genomförts. Till exempel har skolor blivit slumpmässigt uttagna i sådana program i fyra delstater (Killen et al., 1988; Hansen och Graham, 1991; och Perry et al., 1992).

I föränderliga samhällen är det mycket viktigt att förstå hur man minskar psykiska och utbildningsmässiga risker för barn som förflyttas från en skola till en annan. Jason et al. (1992, 1993a, 1993b) riktade in sig på barn som blev sårbara på grund av flytt till nya skolor. Ett projekt randomiserade tio par matchade skolor till ett innovativt behandlingsprogram eller i kontrollgruppen för att avgöra om ett speciellt övergångsprogram fungerade.

Fram till senare delen av 1990-talet var kvalificerade utvärderingar av program för att minska våld i skolorna sällsynta. Bland undantagen noterades en effektstudie av Grossman et al. (1997) av våldsförebyggande kursplaner för årskurs två och årskurs tre. Sex matchande skolor blev slumpmässigt uttagna att använda kursplanen eller fungera som kontrollgrupp. Skillnader i barnens beteende var märkbart och bestod under minst sex månader.

Fram till 1970-talet verkar det som om inget kontrollerat fältförsök hade genomförts i något land för att förstå inverkan av standardiserade prov på studenter. 1975 bestämde sig den irländska republiken att för första gången överväga genomförandet av standardiserade prov i landets grundskolor. Kellaghan, Madaus och Airasian (1982) och deras kollegor på St Patrick's College (Dublin) genomförde en undersökning där 175 lämpliga skolor, matchade och stratifierade, blev randomiserade till olika funktioner. Kontrollgruppen omfattades inte av

standardiserade prov. Interventionen var standardiserade prov, med eller utan återkoppling till lärarna om studenternas prestation.

Randomiserade försök har utförts för att studera utbildningsprogram för förbättring av barnens förståelse av sexuellt högriskbeteende. Till exempel gör Gay (1996) detta i sin avhandling om åtta högstadielklasser, där hälften tilldelades ett nytt rödakorsprogram och hälften utgjorde kontrollgrupp. I Filippinerna riktade Alpasca et al. (1995) också in sig på klassrum inom skolor. I ett omfattande försök i Kalifornien (Kirby et al. 1997a) valdes 102 klassrum slumpmässigt i sex högstadieskolor ut till ett teoridrivet förebyggande program som var mycket beroende av utbildning för jämnåriga ("peer education") för att genomföra programmet. Ett annat Kalifornien-baserat program, Postponing Sex Involvement (PSI) utvärderades genom en komplex forskningsmodell i vilken klassrummen randomiserades med hänsyn till en komponent (Kirby et al., 1997b). Över 50 skolor var involverade.

En annan inriktning av hälsoarbete gäller näringslära. Woodruff (1997), till exempel, beskrev en undersökning i San Diego som omfattade åtta interventionsklasser och nio kontrollklasser som randomiserades till ett nytt program i näringslära från tre kommunala högskolor.

Tidigare exempel att prova olika metoder i olika länder för att förbättra barnens prestation förtjänar erkännande. Sålunda har klassrum i Nicaragua slumpmässigt utvalts för radiobaserad undervisning i matematik och traditionell undervisning för att utröna huruvida det förstnämnda skulle förbättra resultatet i matematik och minska utbildningskostnaderna i relation till det sistnämnda (Dean et al., Jamison, Searle och Suppes, 1980). Hornik et al. (1972) gav en förträfflig beskrivning av ett likadant randomiserat försök i El Salvador som misslyckades. På

1970-talet finansierade det amerikanska utbildningsdepartementet en stor studie om huruvida skolor på ett effektivt sätt kunde använda pengar för att minska isoleringen, orsakad av rasskillnader, och förbättra elevernas resultat. Lämpliga skolor som var frivilliga att delta i undersökningen randomiserades till speciell finansiering och till en kontrollgrupp som inte erhöll speciell finansiering. (Se Coulson, 1978; Reichardt och Rindskopf, 1978 och Wessberg, 1978.)

3.2 Samhällen och geopolitiska enheter som slumpmässigt uttagna enheter

I en studie för att uppmuntra röstberättigade att registrera sig i Chicago, förefaller Gosnell (1927) slumpmässigt ha utvalt vissa grannskap till politiska områden som behandlings- och kontrollgrupper. "Behandlingen" innefattade publicitet, brev och personlig kontakt, ibland på olika språk, i etniskt olika grannskap. Syftet var att ge information om röstregistrering och att uppmuntra till registrering på olika sätt samt testa behandlingen.

Samhällen har varit uttagna enheter i utvärdering av hälsorelaterade program. La Prolle, Bauman och Koch (1992), till exempel, rapporterade om en studie i relativ effektivitet av tre mediekampanjer för förebyggande av rökning bland tonåringar. Forskarna kontrollerade, matchade och randomiserade sedan tio samhällen till en av tre behandlingar och en kontrollgrupp. The Community Intervention Trial for Smoking Cessation, COMMIT, valde ut elva matchade par av lokala samhällen till behandlings- och jämförelsegrupper (Freedman, Green, och Byar, 1990 citerad i Peterson et al., 1992).

I randomiserade försök i fertilitetsinterventioner i fjärran östern, utvaldes samhällen och byar slumpmässigt till olika metoder för att utröna hur barnafödandet kan reduceras (Freedman & Takashita, 1969; Riecken et al., 1974). Ett färre antal lokala samhällen har också använts som

enheter i slumpmässiga studier av riskförebyggande metoder när det gäller HIV (Kelly et al., 1991).

I en mediabaserad röningsförebyggande kampanj utvaldes slumpmässigt storstadsstatistik-områden (SMSAs) till kampanj- respektive kontrollfunktion (Bauman et al., 1991). De federala statistiska myndigheterna specificerade de här områdena på likartat sätt för att klargöra vad ett "storstadsområde" är i motsats till landsbygd. De här områdena används till utformningen av folkräkning och nationella undersökningar. Utbildningsstudier i Cali, Columbia omfattade mycket små geografiska områden i låginkomstgrannskap ("barrios") som slumpmässigt utvaldes till ett kulturberikande och hälsoförbättrande program för förskolebarn i syfte att fastställa dess verkan i jämförelse med slumpmässigt uttagna kontrollområden (McKay et al., 1978).

Några randomiserade försök har utförts därför att integrering av mångsidig service på lokalnivå betraktas som viktig för personer som har psykiska problem och bor i lokalsamhället. Tillgång till Community Care and Effective Service Supports, ACCESS, omfattade åtta städer, var och en innefattande två självständiga jurisdiktioner som slumpmässigt utvaldes till ACCESS eller till kontrollfunktion (Randolph et al., 1997). Cirka 50 organisationer inom varje jurisdiktion samverkade i studien.

Slutligen låt oss betrakta tidigare brottsförebyggande forskning. I Kansas City Patrol Experiment matchades femton polispatrulleringsområden och delades i tre grupper om fem. Därefter jämfördes den relativa inverkan av reaktiv, proaktiv och kontroll (normal) polisbevakning av trakassering (Kelling, Pate, Dieckman och Brown, 1974). Tjugo år senare genomförde Sherman och Weisburd (1995) ett bättre slumpmässigt försök i Minneapolis. Forskarna identifierade över 100 "heta områden", lokala områden med hög kriminalitet, och

valde slumpmässigt ut hälften av områdena till en mer intensiv polispatrullering och hälften till normal patrullering.

3.3 Andra privata och offentliga organisationer som randomiseringsenheter

Program som är utformade för att minska risken för sexuellt överförda sjukdomar kan till exempel vara mer effektiva om programmen är riktade till alla arbetare i fabriker än mot individer som kanske arbetar eller inte arbetar i fabriker. Det är delvis av den anledningen som The National Institute of Allergies and Infectious Diseases i USA har investerat i försök på fabriksbaserad kamratgruppsutbildning (NIAD, 1997). Ingen vet om kamratgruppsutbildningen bland fabriksarbetare kommer att minska infektioner. Projektet involverade cirka 40 fabriker i Zimbabwe, där hälften var slumpmässigt uttagna till program utformat för att minska förekomsten av HIV-infektioner och de återstående för att fungera som kontrollgrupp. Andra randomiserade försök har använt arbetsplatser som enheter i bedömning av program i näringslära och viktkontroll och program för att sluta röka (Simpson et al., 1995).

Ideella serviceorganisationer har tidvis satt in resurser i randomiserade försök. Till exempel, Good Will Industries i USA deltog i kontrollerade undersökningar för att hitta sätt att förbättra ledningen av organisationens butiker (Glaser et al., 1967). I detta fall var självständiga butiker uttagna enheter.

Inom hälsovården deltog ca 40 kommunala sjukhus i Minnesota i ett försök för att undersöka om lokala opinionsbildare inom sjukvården och ett formellt feedbacksystem kunde påverka sjukhusen att införa nya nyttiga terapier för akuta hjärtinfarktspatienter (Soumerai et al., 1998). Teorin bakom programmet är att hela personalens kunskap, inte bara läkarens, tillsammans med uppföljande terapi, är nödvändig för att uppnå förändring. Det var följaktligen inte särskilt bra att endast företa en randomisering av sjukhusläkare. Försökets design innefattade

slumpmässig uttagning av 20 sjukhus till klinisk utbildning och slumpmässig uttagning av 17 sjukhus till kontrollfunktion.

Vårt sista exempel gäller ett program som är utformat att förbättra anställningen av individer som löper högre risk att bli arbetslösa i låginkomstområden i kommuner som behöver ekonomisk vitalisering. I en av varje av sju städer, involverade försöken randomisering av ett kommunalt bostadsområde till programmet och en eller två kommunala bostadsområden som kontrollfunktion. De bakomliggande antagandena till programmets utformning var att förekomsten av lokal medverkan och kollektiva beslut är nödvändig för att förändra lokala samhällen inom bl a utbildning, yrkesträning, arbete och lönenivåer (Riccio, 1998; Bloom, Bos och Lee, 1998).

4. Svårigheter och tänkbara lösningar

Utmaningarna i användning av platser eller andra enheter som uttagna enheter i randomiserade försök är många. Strategier som har uttänkts för att övervinna hinder är värdefulla och diskuteras i det följande.

4.1 Statistisk styrka

Betrakta ett randomiserat fältförsök där två program i läs- och skrivkunnighet jämförs med varandra för att konstatera vilket som är effektivare och billigare. Statistisk styrka hänvisar till vår förmåga att urskilja den relativa effektiviteten av de två läs- och skrivkunnighetsprogrammen. Den här kraften beror på hur läs- och skrivkunnighet mäts. Det beror också på hur många läs- och skrivkunnighetscentra som är slumpmässigt uttagna till ett visst program och på hur många studerande som finns i varje. Den "statistiska styrkan" syftar på vår förmåga att påvisa skillnader i inverkan av interventioner om skillnader verkligen finns.

Hur många centra behövs i den här undersökningen för att säkerställa att dess statistiska kraft är cirka .80? Anta att, som det troligen är, att den egentliga skillnaden mellan programmen är liten (.10) och ställ den statistiska tröskeln (alpha) på .05. Om alla studenter inom skolorna var fristående, skulle cirka 400 studenter testas för varje läroplan för att urskilja inverkan av behandlingen under de här förhållandena.

När likheten bland studenter inom en skola är påtaglig, skulle en större testgrupp behövas för att säkerställa att verkliga skillnader mellan olika interventioner upptäcks. Anta en lägre likhetsgrad (inomklasskorrelation) på .05. Man kan då använda 85 skolor med en testgrupp bestående av 10 studenter i varje, för varje behandling (program) i en formell undersökning. Alternativt kan man använda 44 skolor med 40 studenter i varje.

Enligt La Prolle et al. (1992) var deras undersökning av ett program för samhällsbaserad förebyggande av droganvändning som omfattade en hel stad underdimensionerad. I ett försök var fyra behandlingar utspridda i 10 samhällen. Deras eftertänksamma efter-försöksanalys pekade på att omkring 40 samhällen per grupp skulle behövas för att urskilja en viktig skillnad i effektivitet i röningsförebyggande program.

Platsbaserade randomiserade försök har framgångsrikt varit hänvisade till minst tre metoder för att säkerställa tillräcklig statistisk styrka. Först: enheter som är fristående skall kontrolleras för lämplighet och en rimlig grad av homogenitet. Den andra metoden: enheter skall matchas och sedan randomiseras. Den tredje metoden är implicit: engagera så många enheter som möjligt i försöket.

4.2 Mätssystem och teori

Med en teori om "vad borde hända" menar vi utstakning av på vilket sätt de program som jämförs förväntas engagera och påverka enheterna. Logiken av hur det är förmodat att fungera

måste klargöras. Teorin leder oss att välja vad som på ett mest sofistikerat sätt kan mätas, om och hur bra det kan mätas.

Betrakta Wagenaars et al. (1997) försök som genomfördes på ett flertal platser. Det var utformat för att bedöma om samhällsbaserade program kunde minska minderåriga ungdomars alkoholanvändning. Mobilisering av lokala samhällen betraktades som ett teoretiskt viktigt antagande i utformningen av alkoholpolitik. Lokala samhällens maktstrukturer och studenters och ungdomars attityder observerades. Man gjorde analys av mediabevakningen. Förändringar i den kommunala verksamheten mättes under antagandet att dessa skulle följa den lokala mobiliseringen. Andra satsningar inkluderade en kartläggning av spritbutiker för att utröna om de verkligen underlåtit att kräva legitimation från ungdomar som såg för unga ut. Detta gjordes därför att helt teoretiskt skulle ungdomars minskade tillgång till alkohol resultera i färre alkoholrelaterade trafikolyckor. Det empiriska materialet om alkoholrelaterade trafikolyckor hämtades ur offentlig statistik.

4.3 Att engagera platser och andra enheter

Att engagera platser, administrativa enheter och andra inrättningar i ett randomiserat fältförsök kräver betydande skicklighet. Walker et al. (2000) ger en exceptionellt detaljerad beskrivning av metoder för att värva sjukhus i Storbritannien till randomiserade försök. De riktar uppmärksamheten på att identifiera medarbetarna, informerar dem, ta kontakt med ansvariga för att engagera sjukhuset, förhandla villkor för deltagande, genomföra undersökningen och ge olika slags feedback till ansvariga medverkande. Processen är tidskrävande och utmanande. Enligt vad som kan bedömas av forskarnas framgångar i utförandet av sådana undersökningar är tillvägagångssätten värda allvarligt övervägande.

Betrakta nästa, Ellickson's (1994) uppsats om genomförande av Projekt ALERT, där 30 skolor blev slumpmässigt uttagna till ALERT eller till kontrollfunktion. Syftet med projektet var att undersöka hur bra ALERT fungerade i att förebygga drogmissbruk bland barn och hur länge projektets inverkan varade. För att värva hela skolor till studien, måste man vara medveten om naturliga begränsningar i deras möjlighet att delta. Ellickson (1994) rapporterade att 11 av 60 skolor som erbjöds att delta vägrade att vara med. En skola kunde inte delta på grund av ett domstolsbeslut som krävde stor resurstilldelning till rasjämlighet. Fyra av de elva tackade nej till projektet därför att de redan hade ett interventionsprogram på gång. De andra skolornas vägran grundades på att de saknade resurser till att delta i undersökningen.

4.4 Tidsbestämd och strukturell stabilitet

Vi förväntar oss inte att platser skall ändras mycket över en kort tid. Likväl kan stabiliteten av vissa särdrag hos platser vara låg liksom att deras karaktäristika kan ändra riktning. Bauman et al. (1991), exempelvis, hittade hög positiv korrelation över två årsperioder ($r=.77$) för tonåringars rapporterade senaste rökningstillfälle i ett urval omfattande 10 städer. Forskarna hittade en negativ korrelation ($r=-.31$) för tonåringars provfrekvens i rökning i samma städer. Anledning till detta resultat är oklart. Instabiliteten är klar.

Vanligtvis antar man att platser eller andra enheter som är utvalda till ett program skulle vara strukturellt stabila under pågående undersökning. En skola under ett år förväntas förbli samma skola under en tvåårsperiod. Att döma av erfarenheterna är det klokt att förvänta sig vissa förändringar. Midwestern Prevention Project innebar till exempel att slumpmässigt välja ut skolor till olika funktioner. Pentz (1994) rapporterade att åtta av de 50 högstadieskolorna och gymnasieskolorna som var uppställda som mål i begynnelsen var "stängda eller hade slagits samman med andra skolor under de första tre åren av studien" (s.44). Dessutom förändrades

skolorna på grund av bussning och det faktum att andra skolor tillkom som drog med sig studenter från områden utanför det ursprungliga upptagningsområdet.

Liknande problem påträffades på andra håll. I det nämnda Irish Standardized Testing-experimentet, efter att de hade matchat och slumpmässigt utvalt skolor baserat på folkräkningsdata, kom forskarna på att många viktiga särdrag i skolorna hade förändrats (Kellaghan, Madaus and Airasian, 1982). Delstaten Tennessee's experiment i skolincitament stötte på svårigheter därför att många skolor var stängda eller hade slagits samman med andra skolor (Bickman, 1985). Allt detta skapar komplicerade problem i utformningen av randomiserade försök och deras analys.

4.5 Regionala variationer

För att göra en uppskattning av inverkan av rökningförebyggande program fokuserade Bauman et al. (1991) bara på en region. Trots bemödande om att arbeta i en homogen kontext, blev försöket underdimensionerat. Det innebär att omfattningen av organisationer inom regionen kan ha varit för liten för att utröna den verkliga inverkan av programmen eftersom variationerna inom regionen var för stora. Exempelvis rapporterades av Bauman et al. (1991) att av ungdomar som nyligen rökt var siffrorna 2-7 % för 1985 och 13-20 % för 1987 i tio städer i en region. Förekomsten av rökning under 1987 bland 1985 icke rökare var 3-14 % i de tio städerna.

Stratifiering eller buntning enligt region i ett platsbaserat försök är förnuftig. Men definitioner av en region och implikationer av urvalet har inte utforskats på djupet. En förundersökning inför upprättandet av ett randomiserat försök, pilotförsök och analys av befintlig tid är i alla fall berättigad.

4.6 Obalanserad grupper och begränsad randomisering

Betrakta ett randomiserat försök där ett sampel av samhällen som tilldelas ökade läs- och skrivkunnighetsresurser jämförs med ett urval av samhällen som ännu inte har erhållit dessa resurser. Antal samhällen som är involverade i en sådan studie måste ofta vara relativt litet, mellan 20 till 40 i varje grupp. För analytikern skapar detta ett bekymmer att de två grupper som är slumpmässigt bildade inte kommer att bli "likvärdiga" i början. Det innebär att det kommer att bli en obalans som beror på en slump. Denna "olämpliga slumpmässiga konfiguration" kommer att komplicera jämförelser. En metod som används för att minska problemet i multipla randomiserade fältförsök är *begränsad* randomisering.

I begränsat slumpmässigt urval, betraktas vissa randomiseringskonfigurationer av platser till olika behandlingar som a priori icke-önskvärda. Det innebär att alla möjliga slumpmässiga utvalda konfigurationer under utformning av ett visst försök läggs fram på förhand. De "olämpliga" tas bort. Ett slumpmässigt urval görs sedan bland lämpliga återstående konfigurationer. För forskaren blir en begränsning av slumpmässigheten till lämpliga konfigurationer ett förebyggande av att mycket obalanserade grupper av institutioner blir uttagna till olika programvarianter. Till exempel, Ellickson och Bell (1992) kopplade samman "skolor som var olika och valde slumpmässigt ut dem parvis till försök..." för att åstadkomma balans (s. 85).

Innebörden är att när ett litet antal platser är uttagna som enheter i randomiserade försök, kan vi räkna upp alla möjliga uttagna platser före försöket. Dessutom kan vi eliminera tänkbara uttagningar som är underliga, opassande och så vidare. Efter borttagning av opassande uttagningar, kan vi slumpmässigt välja ut en konfiguration, tilldela institutioner i samstämmighet med detta och utarbeta en rimlig jämförelse av program.

4.7 Genomförande noggrannhet och måttssystem

Det är ingen mening i att bedöma inverkan av ett nytt program utan att kunna verifiera att programaktiviteter inträffar och kan beskrivas. "Genomförandenoggrannhet" avser här i vilken grad ett nytt program faktiskt levereras till avsedda individer. Vi behöver fastsälla att administrativa åtgärder är vidtagna, att informationssystemet fungerar och så vidare. Vetskap om att åtgärder vidtagits är en förutsättning för att kunna utvärdera inverkan.

Försök som syftar till att bedöma interventioner som omfattar "integration" eller "koordination" av tjänster över många organ inom en organisation eller ett lokalt samhälle skapar speciella problem. Att utveckla en sammanhängande definition av integration och mätbar indikator av integration är inte lätt. Betrakta studier av ACCESS inverkan på hemlösa och psykiskt sjuka som exempel. Olika jurisdiktionsenheter kan ha olika uppfattning om och hur de använder sig av olika organ; team mellan olika organ för tillhandahållande av service; ledningssystem mellan olika organ; överenskommelse och avtal mellan olika organ; skapa samarbetsarrangemang; lämplighetsstandard; delad service och lokaler (Randolph et al. , 1997). Att lära sig observera någon av dessa förutsättningar och försäkra deras implementering och mätning är krävande.

5. Sammanfattning: Till vilken effekt?

Inom medicin, kriminologi, socialt arbete, utbildning och andra sektorer har randomiserade försök gett goda bevis för vilka program som fungerar bättre, för vem och för hur lång tid. Försök som innefattar slumpmässigt urval av platser som samhällen, bostadsområden, organisationer, grannskap, skolor och andra enheter för olika interventioner för att erhålla rimlig jämförelse är ännu inte vanliga. Men de kan vara berättigade för teoretiska, statistiska, politiska eller etiska grunder.

De teoretiska grunderna för platsbaserade försök är att program fungerar bättre när de organisatoriska elementen samarbetar, till exempel ett program som omfattar ett helt lokalt samhälle. En grundläggande statistisk utgångspunkt för att koncentrera sig på platser eller institutioner som slumpmässigt uttagna enheter i ett försök är att traditionella statistiska analyser kan vara fel när de baseras på individer istället för institutioner.

Politiska grunder för att rikta in sig på organisationer och andra platser som försöksenheter är att organisationer är omedelbara mål för åtgärder från statliga myndigheter och andra organisationer till skillnad från individer. Etiska och kulturella faktorer kan utgöra grund för slumpmässiga urval av organisationer till alternativa förändringsmodeller, för att främja rimlig jämförelse snarare än slumpmässig uttagning av individer.

Möjligheten att använda platser eller andra inrättningar som enheter i kontrollerade randomiserade försök är uppenbar. Enheter har uttagits slumpmässigt till olika interventioner i försök inom barnbegränsning, förbättringar i socialtjänst, utbildningsreformer, upprätthållande av lag och ordning, program för hälsoriskminskning och så vidare. Grannskap, fabriker, klassrum, skolor, sjukhus, salonger och så vidare har använts som slumpmässigt uttagna enheter.

Självklart finns det svårigheter i genomförandet av sådana försök. Duktiga administratörer, forskare, statstjänstemän och stiftelsemedarbetare har ibland antagit utmaningen. Statistiker och metodologer som förstår utformningen av platsbaserade randomiserade försök kan ibland anpassa utformningen av försöken för att bemöta utmaningarna.

Trots svårigheterna är framtiden för platsbaserade randomiserade försök lovande. De används allt oftare. Platsbaserade försök har utförts inom olika områden såsom utbildning, kriminalitet och brottslighet, psykvård, hälsoriskminskning och socialtjänst. De är viktiga i

framställning av evidens om vilka program som fungerar, för vem de fungerar, om vilka program som inte fungerar, och om vilka program som är lovande.

6. Fotnot

1) Denna artikel är baserad på ett arbete som stöds av amerikanska utbildningsdepartementet, Rockefeller Foundation, Centrum för utvärdering av socialt arbete (CUS) i Stockholm.

2) Referenser i litteraturlistan som är markerad med asterisk (*) rapporterar om försök som omfattar organisationer, grupper eller andra inrättningar som slumpmässigt uttagna enheter i randomiserade försök.

7. Referenser²

- Aplasca, M., Siegel, D., Mandel, J. S., Santana, R., Paul, J., Hudes, E. S., Monzon, O. T, and Hearst, N. (1995) Results of a Model AIDS Prevention Program for High School Students in the Philippines. AIDS, Supplement 1, 7-13. (*)
- Basen-Engquist, K., Parcel, G. S., Harrist, R., Kirby, D., Coyle, K., Banspach, S., and Rugg, D. (1997) The Safer Choices Project: Methodological Issues in School Based Health Promotion Intervention Research. Journal of School Health, 67(9), 365-371. (*)
- Bauman, K. E., LaPrelle, J., Brown, J. D., Koch, G. C. and Padgett, C. A. (1991) The Influence of Three Mass Media Campaigns on Variables Related to Adolescent Cigarette Smoking: Results of a Field Experiment. American Journal of Public Health, 1991, 81, 597-604. (*)
- Bickman, L. (1985) Randomized Field Experiments in Education. New Directions for Program Evaluation, 28, pp. 39-54. (*)
- Bloom, H., Bos, J. and Lee, S. W., (1998) Using Cluster Random Assignment to Measure Program Impacts: Statistical Implications for the Evaluation of Education Programs. New York: New York University, Robert F. Wagner School of Public Service (Research Report). (*)
- Boruch, R. F. (1993(a)) Multi-site Tests in the Civil and Criminal Justice Arena. Invited Presentation, Annual Meeting of the American Society of Criminology (October 30, 1993) Phoenix, Arizona. Available from: Author. University of Pennsylvania, Philadelphia, PA 19104. (*)
- Boruch, R. F. (1993(b)) Multi-site Evaluation and the Children's Initiative. Paper prepared for the Pew Charitable Trusts, Philadelphia, PA. Available from: Author. University of Pennsylvania, Philadelphia, PA 19104. (*)
- Boruch, R. F. (1997) Randomized Experiments for Planning and Evaluation: A Practical Guide. Thousand Oaks, CA: Sage.
- Boruch, R. F. and Foley, E. (2000) The Honestly Experimental Society: Sites and Other Entities as the Units of Allocation an Analysis in Randomized Trials. In L. Bickman (Ed.) Validity and Experimentation: Donald Campbell's Legacy Volume 1. Thousand Oaks, CA, London, New Delhi: Age Publications.
- Botvin, G. J., Baker, E., Dusenburg, L., Botvin, E. M., and Diaz, T. (1995) Long Term Follow-up Results of a Randomized Drug-Abuse Prevention Trial in a white Middle Class Population. Journal of the American Medical Association, 273, 1106-1112. (*)

- Campbell, D. T. (1969) Reforms as Experiments. American Psychologist, 24(4), 408-429. (*)
- Campbell, D. T. (1988) The Experimenting Society. Chapter 11 in S. Overman (Ed.) Methodology and Epistemology for Social Science: Selected Papers by Donald T. Campbell. Chicago: University of Chicago Press, pp. 290-314.
- Campbell, D. T. and Stanley, J. C. (1963) Experimental and Quasi-experimental Designs for Research Teaching. In N. L. Gage (Ed) Handbook of Research on Teaching. Chicago, IL: Rand McNally, pp 171-246.
- Coulson, J. E. (1978) National Evaluation of the Emergency School Aid Act (ESAA): A Review of Methodological Issues. Journal of Educational Statistics, 3(3), 1-60. (*)
- Coyle, S. L., Boruch, R. F., and Turner, C. F. (Eds.) (1991) Evaluating AIDS Prevention Programs (Expanded Edition). Washington, DC: National Academy of Sciences. (*)
- Coyle, K, Kirby, D., Purcel, G., Basen-Engquist, K., Banspach, S, Rugg, D., and Well, M. (1996) Safer Choices: A Multicomponent School Based HIV/STD and Pregnancy Prevention Program for Adolescents. Journal of School Health, 66(3), 89-84. (*)
- Dean, J., Seare, B., Galda, K., and Heyneman S. P. (1981) Improving Elementary Mathematics Education in Nicaragua: An Experimental study of the Impact of Textbooks and Radio on Achievement. Journal of Education Psychology, 73(4), 556-567. (*)
- Donner, A. and Klar, N. (2000) Design and Analysis of Cluster Randomized Trials in Health Research. New York: Oxford University Press.
- Ellickson, P. L. (1994) Getting and Keeping Schools and Kids for Evaluation Studies. Journal of Community Psychology (Monograph Series: CSAP Special Issue), pp. 102-116. (*)
- Ellickson, P. L. & Bell, R. M. (1992) Challenges to Social Experiments: A Drug Prevention Example. Journal of Research in Crime and Delinquency, 29(1), pp. 79-101. (*)
- Ellickson, P. L. & Bell, R. M. (1990) Drug Prevention in Junior High: A Multi-site Longitudinal Test. Science, 247, pp. 1299-1305. (*)
- Ennett, S. T., Tobler, N. S., Ringwalt, C. L., and Flewelling, R. L. (1994) How effective is Drug Abuse Resistance Education? A Meta-analysis of Project DARE's Outcome Evaluations. American Journal of Public Health, 84(9), 1394-1401. (*)
- Fairweather, G. W., Sanders, D. H., & Tornatzky, L. G. (1974) Creating Change in Mental Health Organizations. New York: Pergamon. (*)
- Fairweather, G. W. and Tornatzky, L. G. (1977) Experimental Methods for Social Policy Research. New York: Pergamon Press. (*)

- Farrington, D. P. (1997) Evaluating a Community Crime Prevention Program. Evaluation, 3. (*)
- Flay, B. R., Ryan, K. B., Best, J. A., Brown, K. S., Kersell, M. W., d'Avernas, J. R. & Zanna, M. P. (1985) Are Social-psychological Smoking Prevention Programs Effective? The Waterloo Study. Journal of Behavioral Medicine, 8(1), pp. 37-59. (*)
- Freedman, R. & Takashita, J. T. (1969) Family Planning in Taiwan: An Experiment in Social Change. Princeton, NJ: Princeton University Press. (*)
- Gay, K. E. M. (1996) Collaborative School-based Research: The Creation and Implementation of an HIV/AIDS Prevention Curriculum for Middle School Students. PhD Dissertation, University of Pennsylvania, Philadelphia, PA. (*)
- Glaser, E. M., Coffey, H. A., and others (1967) Utilization of Applicable Research and Demonstration Results. Los Angeles, CA: Human Interaction Research Institute. (*)
- Gosnell, H. F. (1927) Getting Out the Vote: An Experiment in the Stimulation of Voting. Chicago: University of Chicago Press. (*)
- Green, S. B. and Washington, A. E. (1991) Evaluation of Behavioral Interventions for Prevention and Control of Sexually Transmitted Diseases. In: J. N. Wasserheit, S. O., Aral, K. K., Holmes, and P. J. Hitchcock (Eds.) Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era. Washington, D.C.: American Society for Microbiology, pp. 345-352.
- Grossman, D. C., Neckerman, H. J., Koepsall, T. D., Liu, P., Asher, K. N., Beland, K., Frey, K., and Rivara, F. P. (1997) Effectiveness of a Violence Prevention Curriculum among Children in Elementary School: A Randomized Controlled Trial. Journal of the American Medical Association, 277(20), 1605-1611, (*)
- Hornik, R. (1991) Alternative Models of Behavior Change. In J. N. Wasserheit, S. O., Aral, K. K., Holmes, and P. J. Hitchcock (Eds.) Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era. Washington, D.C.: American Society for Microbiology, pp. 201-218.
- Hornik, R. C., Ingle, H.T., Mayo, J. K., McAnany, E. G., and Schramm, W. (1972) Television and Education Reform in El Salvador. (Report No. 14) Stanford University, Institute for Communication Research. (*)
- Jamison, D., Searle, B., & Suppes, P. (1980) Radio Mathematics in Nicaragua. Stanford, CA: Stanford University Press. (*)
- Jason, L. A., Weine, A. M., Johnson, J. H., Donner, K. E., Kuraski, K. S., & Sohlberg, L. (1993a) Journal of Emotional and Behavioral Disorders, 1(1), pp. 65-70. (*)

- Jason, L. A., Weine, A. M., Johnson, J. H., Sohlberg, Filippelli, Turner, E., & Lardon, C. (1992) Helping Transfer Students: Strategies for Educational and Social Readjustment. San Francisco: Jossey-Bass. (*)
- Jason, L., Johnson, J. H., Danner, K. E., Taylor, S., and Krasaki, K. S. (1993b) A Comprehensive, Preventive, Parent-Based Intervention for High Risk Transfer Students. Prevention in Human Services, 10(2), 27-37. (*)
- Kaftarian, S. J. & Hansen, W. B. (1994) (Eds.) Community Partnership Program: Center for Substance Abuse Prevention. CSAP Special Issue/Monograph Series. Journal of Community Psychology. (*)
- Kellaghan, T., Madaus, G. F., Airasian, P. W. (1982) The Effects of Standardized Testing. Boston/The Hague/London: Kluwer-Nijhoff. (*)
- Kelling, G. L., Pate, T., Dieckman, D., & Brown, C. E. (1974) The Kansas City Preventive Patrol Experiment: A Summary Report. Washington, DC: Police Foundation. (*)
- Kelly, J.A., Lawrence, J.S., Diaz, Y. E. and others. (1991) HIV Risk Behavior reduction Following Intervention with Key Opinion Leaders: An Experimental Analysis. American Journal of Public Health,81, 168-171. (*)
- Killen, J. D., Telch, M. J., Robinson, T. N., Maccoby, N., Taylor, C., & Farquar, J. W. (1988) Cardiovascular Disease Risk Reduction for Tenth Graders: A Multiple Factor School-based Approach. Journal of the American Medical Association, 260(12), pp. 1728-1733. (*)
- Kirby, D., Korpi, M., Adivi, C. and Weismann, J. (1997a) An Impact Evaluation of Project SNAPP: An AIDs Prevention and Pregnancy Middle School Program. AIDS Education and Prevention, 9 (Supplement A), 44-61. (*)
- Kirby, D., Korpi, M, Barth, R. P., and Cagampang, H. H.(1997b) The Impact of Postponing Sexual Involvement Curriculum among Youths in California. Family Planning Perspectives, 29, 100-108. (*)
- Kuusi, Pekka (1957) (WestPhaler, A. Translator). Alcohol Sales in Rural Finland. Volume 3 Publication of the Finish Foundation for Alcohol Studies. Stockholm, Sweden: Alm of Vist and Wiksell.
- LaPrelle, J., Bauman, K. E. & Koch, G. G. (1992) High intercommunity variation in adolescent cigarette smoking in a 10-community field experiment. Evaluation Review, 16(2), pp. 115-130. (*)
- Leviton, L., Valdiserri, R., Lyter, D., Callahan, C., Kingsley, L., Huggins, J., and Rinalde, C. R. (1990) Preventing HIV Infection in Gay and Bisexual Men: Experimental Evaluation of

- Attitudes Changes from Two Risk Reduction Experiments. AIDS Education and Prevention, 2(2), 95-108. (*)
- Lorian, R. P. (1994) Epilogue: Evaluating the Community Partnership Program. Reflections on a Name. Journal of Community Psychology (Monograph Series: CSAP Special Issues), pp. 201-205. (*)
- McKay, H., McKay, A., Sinnestera, L., Gomez, H., and Lloreda, P. (1978) Improving Cognitive Ability in Chronically Deprived Children. Science, 200(4), 270-278. (*)
- Moskowitz, J. et al. (1984) The Effects of Drug Education and Follow-up. Journal of Alcohol and Drug Education, 3, pp. 45-49. (*)
- Murray, D. (1998) Design and Analysis of Group Randomized Trials. Oxford and New York: Oxford University Press.
- Murray, D. M., McKinlay, S. M., Martin, D., Donner, A. P., Dwyer, J. H., Raudenbush, S. W., & Graubard, B. I. (1994). Design and Analysis Issues in Community Trials. Evaluation Review, 18(4), pp. 493-514. (*)
- Murray, D. M. and Wolfinger, R. D. (1994) Analysis Issues in the Evaluation of Community Trials: Progress Toward Solutions in SAS/STAT Mixed. Journal of Community Psychology (Monograph Series: CSAP Special Issue), pp. 140-154. (*)
- Murray, D., Moskowitz, J. M., and Dent, C. W. (1996) Design and Analysis Issues in Community-Based Drug Abuse Prevention. American Behavioral Scientist, 39(7), 853-867. (*)
- Pentz, M. A. (1994) Adaptive Evaluation Strategies for Estimating the Effects of Community Based Drug Abuse Prevention Programs. Journal of Community Psychology (Monograph Series CSAP Special Issue), pp. 5-25. (*)
- Perry, C., Parcel, G. S., Stone, E., Nader, P., McKinlay, S. M., Leupker, R. V., and Webber, L. S. (1992) The Child and Adolescent Trial for Cardiovascular Health (CATCH): An Overview of Intervention Program and Evaluation Methods. Cardiovascular Risk Factors, 2(1), pp. 36-43. (*)
- Peterson, P. L., Hawkins, J. D., & Catalano, R. F. (1992) Evaluating Comprehensive Community Drug Risk Reduction Interventions. Evaluation Review, 16(6), pp. 579-602. (*)
- Randolph, F, Basinsky, M., Leginski, W., Parker, L., and Goldman, H. H. (1997) Creating Integrated Service Systems for Homeless Persons with Mental Illness: The Access Program. Psychiatric Services, 48(3), 369-373. (*)
- Reichardt, C. S. & Rindskopf, D. (1978) Randomization and Educational Evaluation: The ESAA Evaluation. Journal of Educational Statistics, 3(1), 61-68. (*)

- Reiss, A. J. & Roth, J. A. (Eds.) (1993) Understanding and Preventing Violence. Washington, DC: National Academy of Sciences Press.
- Riccio, J. A. (1998) A Research Framework for Evaluating Jobs-Plus, A Saturation and Place-Based Employment Initiative for Public Housing Residents (Working Paper). New York, Manpower Demonstration Research Corporation. (*)
- Riecken, H. W., Boruch, R. F., Campbell, D. T., Caplan, N., Glennan, T. C., Pratt, J. W., Rees, A., & Williams, W. (1974) Social Experimentation: A Method for Planning and Evaluating Social Programs. New York: Academic Press. (*)
- Rosenbaum, D. P., Ringwalt, C., Curtin, T. R., Wilkinson, D., Davis, B., & Taranowski, C. (1991) Second Year Evaluation of D.A.R.E. in Illinois. (Available from: D. P. Rosenbaum Center for Research in Law and Justice, University of Illinois at Chicago, Chicago, Illinois 60607). (*)
- Schaps, E., Moskowitz, J., Condon, J., & Malvin, J. (1982) A Process and Outcome Evaluation of a Drug Education Course. Journal of Drug Education, 12, pp. 245-454. (*)
- Shadish, W. R., Cook, T. D., and Campbell, D.T. (2002) Experimental and Quasi-experimental Designs for generalized Causal Inference. New York: Houghton Mifflin.
- Sherman, L. and Weisburd, D. (1995) General Deterrent Effects of Police Patrol in Crime “Hot Spots”: A Randomized Controlled Trial. Justice Quarterly, 12(40), 625-648. (*)
- Simpson, J. M., Klar, N. and Donner, A. (1995) Accounting for Cluster Randomization: A Review of Primary Prevention Trials, 1990 through 1993. American Journal of Public Health, 85(10), 1378-1383. (*)
- Soumerai, S. B., McLaughlin, T. J., Gurwitz, J. H., Guadagnoli, E., Hauptman, P. J., Borbas, C., Morris, N., McLaughlin, B., Gao, X., Willison, D. J., Asinger, R. and Gobel, F. (1998) Effect of Local Medical Opinion Leaders on Quality of Care for Acute Myocardial Infarction. Journal of the American Medical Association, 279(17), 1358-1363. (*)
- Soydan, H. (1998) (Issue Editor) Evaluation Research and Social Work. Scandinavian Journal of Social Work Welfare 7 (2).
- Wagenaar A. C., Murray, D. M., Wolfson, M., Forster, J. L., & Finnegan, J. R. (1994) Communities mobilizing for Change on Alcohol: Design of a Randomized Community Trial. Journal of Community Psychology (Monograph Series/CSAP Special Issue), pp. 79-101. (*)
- Wagenaar, A. C., Murray, D. M., Gehan, J. P., Wolfson, M., Forster, J. L., Toomey, T. L., Perry, C. L., and Jones-Webb, R. (1997) Communities Mobilizing for Change on Alcohol (CMCA): Outcomes from a Randomized Trial. Report. University of Minnesota. Submitted. (*)

- Walker, A. E., Campbell, M. K., Grimshaw, J. M., and the TEMPEST Group (2000) A Recruitment Strategy for Cluster Randomized Trials in Secondary Care settings. Journal of Evaluation in Clinical Care Settings, 6(2), 185-192.
- Wasserheit, J. N., Aral, S. O., Holmes, K. K., and Hitchcock, P. J. (Eds.) (1991) Research Issues in Human Behavior and Sexually Transmitted Diseases in the AIDS Era. Washington, D.C.: American Society for Microbiology. (*)
- Weisberg, H. (1978) How Much does ESAA Really Accelerate Academic Growth. Journal of Educational Statistics, 3(1), 69-78. (*)
- Weisburd, D., Sherman, L., & Petrosino, A. J. (1990) Registry of Randomized Criminal Justice Experiments in Sanctions. Washington, DC: National Criminal Justice Reference Service (SRO 19000-00/129725).
- Woodruff, S. I. (1997) Random Effects Models for Analyzing Clustered Data from a Nutrition Education Intervention. Evaluation Review, 21(6), 688-697. (*)