

Om och varför? Den potentiella nyttan av att inkludera programteoriutvärderingar i metaanalys.

anthony petrosino

Två envisa utmaningar för att förstå effekterna av socialpolitiken är ansamlingen av rön från tidigare utvärderingsstudier och att förstå varför ett program lyckas eller misslyckas. Under de senaste trettio åren har metaanalys och programteoriutvärdering lanserats som metoder för att svara på denna utmaning. Författaren visar hur till och med enkla programteoriutvärderingar kan komma till nytta om de används vid metaanalys.

Inledning

Att bestämma om och varför ett program fungerade är en riskabel sysselsättning. Ett slumpmässigt experiment – om det realiserar och fullföljs med totalt oberoende – kan ge det minst tvetydiga svaret på frågan »fungerade programmet?« Men om man inte planerar och analyserar ytterligare data förutom resultatmätningarna kan ett experiment sällan ge svar på frågan »varför fungerade interventionen?«¹ Även om vi

Anthony Petrosino is Research Fellow at the Center for Evaluation, Initiatives for Children Program of the American Academy of Arts & Sciences. He is also a Research Associate and Project Manager for the Study of Decisions in Education at Harvard University

skulle lyckas med att utforma en studie som svarar på både »om« och »varför«, vilket inte vore en liten bedrift i sig, finns det ytterligare ett problem. Enstaka, fristående utvärderingar kommer sällan att ge definitiva resultat. Resultaten från utvärderingar av samma program varierar ofta från fall till fall beroende på skillnader i kunder, personal och så vidare (Lipsey, 1997).

Två olika perspektiv har emellertid framträtt som tillsammans skulle kunna ge bättre svar på »om« och »varför«. Det första är programteori. Under de senaste trettio åren har utvärderare skrivit övertygande –

¹ Vissa människor vill kanske hävda att faktoriella experiment isolerar mekanismerna till varför ett program fungerar.

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

och ofta – om behovet av att oförbehållsamt pröva programteori vid utvärdering (t.ex. Bickman, 1987; Chen and Rossi, 1992; Weiss, 1972). Även om vokabulären skiftar mellan dessa författare är alla överens om att utvärderingar, när det är möjligt, bör tydligt formulera och pröva de underliggande antagandena om varför programmet borde »fungera«. ² I konsekvens med andra artiklar i den här volymen, har jag använt programteoriutvärdering (program-theory evaluation – PTE) för att beskriva studier som använder sådana termer som teori-baserad eller teoristyrd för att beskriva liknande metoder. ³

Intresset för PTE går hand i hand med framkomsten av en »vetenskap om litteraturöversikter«. Även om problemet med att tolka skilda men liknande studier har uppmärksamats ända sedan 1904 sköt inte vetenskapen om litteraturöversikter fart förrän på 1970-talet med metaanalysen (Hunt, 1997). Eftersom man insåg att de traditionella metoderna för att sammanställa forskningsresultat var bristfälliga framträdde metaanalysen som en ytterst noggrann metod för sammanfattning av resultat från tidigare forskning (t.ex. Lipsey, 1990). Till skillnad mot PTE har användningen av metaanalys varit utbredd. Lipsey och Wilson (1993) lyckades identifiera drygt 300 metaanalyser av behandlingar inom det sociala fältet och inom utbildningsområdet.

2 Observera att det finns avvikande synpunkter beträffande värdet av PTE (t.ex. Stufflebeam, 2001).

3 Jag föredrar termen «orsaksmodell» på grund av den allmänna förvirring som råder kring termen «teori» (Petrosino, 2000).

Författare som Cordray (1992) och Lipsey (1997) menar att kombinationen av PTE och metaanalys skulle kunna ha fördelar. Lipsey (1997) visade till exempel hur de skulle kunna användas tillsammans för att skapa teorier om social intervention. I den här artikeln bygger jag på dessa tidigare argument för att visa hur upplagringen av kunskap från PTE genom metaanalys skulle kunna ge nyttiga data för socialpolitiska beslut och beslut inom den sociala praktiken. Både hypotetiska illustrationer och faktiska exempel kommer att användas.

Utvärderingsmodell i ett steg

Många utvärderingar prövar bara effekten av en intervention på resultatmätningarna – vilket ibland kallas ettstegsmodeller

Tack

En äldre version av denna artikel har tidigare publicerats som «Om och varför: Den potentiella ömsesidiga nyttan av att inkludera programteoriutvärderingar i metaanalys» i *New Directions in Evaluation*, 87, 59-70 ©2000, Jossey-Bass, ett dotterföretag till John Wiley & Sons, Inc. Tidigare utkast har framlagts för Harvard Children's Initiative Evaluation Task Force (april 1998) och American Evaluation Association (november 1998). En docentur vid Harvard Children's Initiative finansierad med medel från Spencer Foundation samt ett stipendium från Mellon Foundation till Center for Evaluation underlättade detta arbete. Jag är tacksam för synpunkter på tidigare utkast av denna artikel från Mary Askew, Anne Barten, Len Bickman, Iain Chalmers, Jodi Delibertis, Tim Hacsí, Tracy Huebner, Mark Lipsey, Heather McMillan, Frederick Mosteller, Pamela Perry, Carolyn Petrosino, Patricia Rogers, Sean Riordan, Robert Rosenthal, Haluk Soydan, Carol Weiss, David Wilson och Stuart Yeh. Det är emellertid endast författarens åsikter som det ges uttryck för i den här artikeln.

(Weiss, 1997). Problemet med ettstegsmodellen är att den inte förklarar varför ett program skulle påverka resultatet (Chen och Rossi, 1992). Den behandlar inte den orsaksmässiga komplexiteten som finns i många program som riktar in sig på sådana resultat som till exempel kriminellt beteende. Vissa sociala interventioner fungerar genom indirekta processer: en behandling inleds i ett sammanhang och förväntas ta andra avgörande mekanismer i anspråk för att påverka resultatet i ett annat sammanhang (Donaldson, 2001). Skolbaserad drogprevention inriktar sig till exempel på drogmissbruk bland ungdomar, vilket förmodas ta i anspråk till exempel sådana mekanismer som motstånd från kamraterna. Dessa mekanismer förmodas öka motståndet mot att använda droger både i och utanför skolan. Ettstegsmodellen skulle inte ta hänsyn till mätningen av sådana underliggande mekanismer som kamratmotstånd utan bara inrikta sig på sådana resultat som droganvändning.

Hur utvärderingar i ett steg påverkar metaanalysen

Valet att utföra ettstegsutvärderingar har konsekvenser för forskningssyntesen eftersom det är de ursprungliga studierna som förser metaanalysen med data. Det är viktigt att utvärderare inser att det sätt på vilket de genomför och redovisar sina studier påverkar senare genomgångar.

I de flesta metaanalyser tar granskarna fram en effektstorlek för att uttrycka programmets inverkan på den aktuella resultatmätningen. I en utvärdering i Illi-

nois, till exempel, utvaldes 94 ungdomsbrottslingar slumpmässigt för att delta i ett fängelsebesök, medan 67 ungdomar inte kontaktades alls (Greater Egypt Regional Planning & Development Board, 1979). I programmet ingick också ett tillfälle då ungdomarna fick möjlighet att prata med fångarna, som återgav realistiska berättelser om fängeslivet, inklusive våldtäkter och mord. Programmet, av typen »Scared Straight« (Skrämd till skötsamhet), hade för avsikt att avskräcka ungdomsbrottslingarna från vidare brottslighet. Efter sex månader fann utvärderarna att 17 % av experimentgruppen hade anhållits på nytt jämfört med kontrollgruppens 12 %, vilket ger en effektstorlek på -0,14.⁴ Minustecknet beror på att programmets påverkan gick i motsatt riktning mot den förväntade.

I de flesta metaanalyser anger man genomsnittet av sådana här effektstorlekar för alla inkluderade studier (t.ex. »program som avser att minska ungdomsbrottsligheten hade en genomsnittlig effektstorlek på 0,10«). I vissa metaanalyser anges genomsnittet av effektstorlekarna för specifika behandlingar som använder vida beteckningar (t.ex. »yrkesutbildningsprogram för att minska ungdomsbrottsligheten hade en genomsnittlig effektstorlek på 0,05«). En sådan beteckning tjänar två syften: den tillhandahåller en rubrik som beskriver den grundläggande interventionen, men är så pass vid att den innefattar mer än bara ett fåtal studier. Dessa kategorier har konsekvenser för metaanalysen. Kognitiv terapi, beteendeterapi, individuella terapeutiska

4 I Lipsey (1990) finns omvandlingsformler för effektstorlekar.

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

samtal och gruppterapi är samtliga behandlingar för sexualförbrytare (Laws, 1989). Om utvärderingarna av dessa behandlingar hade redovisats, skulle metaanalysen kunna ta vid och effektstorlekarna beräknas. De flesta metaanalytiker sätter som behörighetskriterium att utvärderingarna skall ha med antingen en kontrollgrupp eller en jämförelsegrupp.

I tabell 1 visas de hypotetiska resultaten från en metaanalys av studier kring behandling av sexualförbrytare. Av tabellen framgår att effektstorleken för kognitiv behandling var 0,30. Skulle ett sådant rön kunna vara viktigt? Rosenthal och Rubin (1982) har utarbetat en metod att översätta effektstorlekar till skillnader i procent, mera känt som Binomial Effect Size Display (BESD). Genom att använda BESD kan man översätta effektstorleken 0,30 till en 15-procentig förbättring för den kognitiva gruppen jämfört med kontrollgruppen. En genomsnittlig minskning med 15 procent av återfallen för sexualförbrytare skulle vara viktig.⁵

Men den vida beteckningen »kognitiv behandling« döljer viktig information om programmen i denna kategori. Det finns många olika slags kognitiva program för sexualförbrytare med stor variation i hur bara en enda typ av program genomförs på olika håll. Ansvariga för finansieringen eller

5 Rosenthal och Rubins (1982) BESD kan omvandla effektstorlekar till skillnader i procent, vilket gör att även lekmän kan ta till sig resultaten. Enkelt uttryckt är BESD lika med hälften av effektstorleken d (om man t.ex. vid $d=0,30$ förutsätter en återfallsprocent (baseline) på 50 %, återfaller experimentgruppen i 32,5 % av fallen och kontrollgruppen i 47,5 %).

Tabell 1.
Hypotetisk metaanalys av ettstegsutvärderingar, behandling av sexualförbrytare (ordnade efter effektstorlek)

Program (N)	Effekt på återfall BESD i sexualbrott	
kognitivbaserat (11)	0,30	+15 %
beteende (10)	0,11	+5,5 %
individuellt (12)	0,01	+0,05 %
grupp (8)	-0,05	-2,5 %

införandet av kognitiva program anser dock kanske inte att resultaten i tabell 1 hjälper dem i deras beslutsfattande. Det finns många program som anses vara kognitiva och den vida beteckningen antyder inte vilket av de kognitiva programmen de bör använda.

Hur PTE skulle kunna användas vid metaanalys

Ett sätt att komma ifrån det här problemet är att öka antalet rigorösa utvärderingar med hjälp av programteorimetoden. I slutet av den här artikeln finns ett praktiskt förslag på hur man gör det. Till och med enkla PTE som bara inriktar sig på en enda nyckelmekanism och resultatet skulle visa att programmen fungerar genom en eller annan mekanism. Metaanalytiska rön skulle då kunna kategoriseras genom den nyckelmekanism som prövades i de ursprungliga utvärderingarna. Om sådana mekanismer prövades i ett antal PTE skulle metaanalysen bli bättre lämpad att ge vägledning om vad som är en effektiv intervention.

För att återgå till tabell 1, så visade rönen

att kognitiva program var den effektivaste strategin för att minska återfallen. Men det gavs ingen information om varför de kognitiva programmen var mer effektiva. Ännu mer problematiskt är det att den genomsnittliga effektstorleken innefattar program med olika effekter; vissa var antagligen mycket effektiva, medan andra förmodligen hade mindre effekt på återfallsrisken än till och med icke-kognitiva behandlingar (t.ex. grupperapi).

I stället för de elva utvärderingarna av kognitiv behandling som redogjordes för i tabell 1, vad skulle hända om ett större antal enkla PTE inkluderades i metaanalysen? En enkel PTE är vad Lipsey och Pollard (1989) beskriver som en tvåstegsmodell: mätning och prövning av minst en förändringsmekanism samt ett resultat. Om man tillämpade en sådan modell på utvärdering av behandlingsprogram för sexualförbrytare skulle modellen pröva huruvida programmet först förändrade någonting som sedan i sin tur påverkade återfallsrisken.

I tabell 2 jämförs resultaten från tabell 1 med en hypotetisk metaanalys av en rad enkla PTE. Fördelen med PTE-metaana-

lysen är att den ger ledtrådar om förändringsmekanismer. Effektiva program känns lättare igen genom den nyckelmekanism de tar i anspråk. I tabell 2, till exempel, är kognitiva program som ökar färdigheten hos sexualförbrytare att identifiera och minska sina egna högrisksituationer mer effektiva när det gäller att minska återfallsrisken. Om kognitiva program skall användas är de till och med mer effektiva om de används samtidigt som man inriktar sig på empatin hos brottslingen. Sådana rön skulle vara mycket användbara när det gäller att ge vägledning inför ett beslut.

Wilson och hans kollegor (2000) har givit ett tidigt och äkta exempel på hur en sådan metaanalys kan gå till i praktiken. I sin studie analyserade de 33 studier som utvärderade effekten av arbete och yrkesutbildningsprogram på senare återfall från förövarnas sida. Bara nio studier redogjorde för effekten av ett program på både återfallsrisken och på den grundläggande nyckelmekanismen: sysselsättning. Med andra ord: Är det så att de program som ökar sysselsättningen (mekanismen) även minskar återfallsrisken? Trots att interna validitets-

Tabell 2.
Jämförelse mellan hypotetisk metaanalys av ettstegsutvärderingar och hypotetisk metaanalys av PTE

Med ettstegsutvärderingar:		Med enkla PTE:		
Program (N)	Återfall i sexualbrott	Kognitivbaserade program (N)	Effekt på mekanism	Återfall i sexualbrott
kognitivbaserade (11)	0,30	färdighet, empati med offer (7)	0,61	0,44
beteende (10)	0,11	färdighet i högrisksituationer (14)	0,55	0,38
individuella (12)	0,01	minska rationalisering (8)	0,28	0,12
grupp (8)	-0,05	öka empatin med offren (12)	0,25	0,09

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

hot var vanliga i de studier som Wilson et al. (2000) tog exempel ifrån rapporterar de att »... större skillnader i återfallsfrekvens hos programdeltagare jämfört med sådana som inte deltog står i samband med större skillnader i sysselsättningsstatus vid uppföljning. I program där man konstaterar en påverkan på sysselsättningen brukar också noteras en påverkan på återfallsrisken« (s. 361).

Minimikrav på PTE för metaanalys

Om PTE skall kunna användas vid metaanalys bör de leva upp till tre kriterier: (1) en tydlig orsakmodell för hur programmet kommer att påverka resultaten, (2) prövning av minst en underliggande mekanism som interventionsvariabel tillsammans med resultaten och (3) kontrollgruppsdata redovisade för båda variablerna. Det första kriteriet kräver att en prospektiv och tydlig modell prövas vid utvärderingen. Förutom att tydligheten minskar antalet gissningar som läsaren måste ägna sig åt beträffande vilken programteori det rörde sig om, så förhindrar det prospektiva kravet att data efterhand anpassas till modellen.⁶

Det andra kriteriet kräver att PTE innefattar minst en mellanliggande variabel.

⁶ Vissa utvärderingar samlar ihop ansevärd mängder data om deltagarna och en del information skulle kunna conceptualiseras som mellanliggande variabler. Metaanalysen kan hantera sådana «vardagsutvärderingar» – om det finns tillräckligt många – och undersöka korrelationerna för både de mellanliggande variablerna och resultatmätningarna.

Inom utvärderingsvärlden är en mellanliggande variabel något som programmet måste påverka eller förändra för att på ett positivt sätt påverka slutresultatet (Lipsey och Pollard, 1989). Vissa utvärderingar där programteori har ingått har inte försökt pröva några nyckelförbindelser eller -mekanismer i modellen (Petrosino, 2000). Om det bara är resultatdata som analyseras och redovisas spelar det ingen roll om en programteori är tydlig för då kommer inte sådana utvärderingar att ge mer information än ettstegsmodeller – programmets orsakmodell prövades inte. Enligt det tredje kriteriet måste kontroll- och jämförelsedata redovisas för både de mellanliggande variablerna och resultatvariablerna. Om data om den mellanliggande variabeln bara redovisas beträffande behandlingsgruppen ger utvärderingen inte mycket belägg för att effekten på den mellanliggande variabeln skulle ha inträffat utan programmet (Cook, 2000).

Utvärderingar av processresultat motsvarar inte heller minimikraven för PTE eftersom de inte ger några data om de underliggande mekanismerna. Till och med när utvärderare kopplar samman processdata med resultaten i sina analyser återspeglar de programaktiviteternas påverkan och graden av trohet gentemot resultatet – inte den underliggande förändringsteorin (Weiss, 1997). Det är något med de här aktiviteterna som borde ta en avgörande mekanism i anspråk. Vad är detta något? Det är det som PTE måste ge uttryck för och pröva (se Weiss, 2000).

Hur skulle metaanalys av PTE kunna vara till ledning för mer omfattande sociala teorier?

Inom PTE är teori en förklaring till hur programmet kommer att åstadkomma de avsedda resultaten. Mer omfattande vetenskapliga teorier erbjuder generella förklaringar till sådana fenomen som kriminalitet, dålig inlärning och till och med om hur program genomförs. En metaanalys av PTE skulle potentiellt kunna vara till ledning för sådana mer omfattande teorier.

I Chandlers (1973) experimentella utvärdering av en förebildsintervention med brottsbenägna ungdomar prövade han till exempel en tvåstegsmodell: en minskning av självupptagenheten (dvs. bristen på empati med andra) skulle minska brottsligheten (Lipsey och Pollard, 1989). Chandler genomförde en uppföljning efter två år och kom fram till att behandlingen åstadkom statistiskt signifikanta minskningar i både självupptagenhet och brottslighet.

I stället för att bara visa ett experiment ger tabell 3 ett hypotetiskt exempel på hur 50 studier som Chandlers skulle kunna vara till ledning för teorier om brottslighet. I tabell 3 prövar tio PTE självupptagenhets-

modellen och fyra andra grupper av PTE prövar olika förändringsmekanismer. Tabellen visar hypotetiska effektstorlekar för var och en av fem kategorier. Intervention som inriktade sig på självupptagenheten åstadkom hypotetiskt sett större effekter på både den mellanliggande variabeln och den påföljande brottsligheten. Ett sådant rön antyder att självupptagenhet är en avgörande länk i uppkomsten och utvecklingen av brottslighet.

De hypotetiska rönen visar också att interventioner i allmänhet hade mindre effekt på mätningar av självkänsla, yrkesskicklighet, familjefunktion och rädsla för påföljder. Sådana rön skulle kunna få brottsteoretiker att på nytt undersöka förhållandet mellan sådana faktorer och den påföljande brottsligheten. Naturligtvis skulle effekterna kunna bero på att programmet genomfördes dåligt eller på en allmänt ineffektiv behandling (om de tio studierna grundades på en gemensam behandlingstyp). Men när allt kommer omkring skulle en generation av PTE för metaanalys kunna ge en del teorier om ledtrådar, inte minst om sådana problem som brottslighet.

Tabell 3.

Hypotetiska effektstorlekar för mellanliggande variabler och resultatvariabler (ordnade efter effekt på den mellanliggande variabeln)

Mellanliggande (N) variabel	Effekt på mellanliggande variabel	Effekt på brottslighet
självupptagenhet (10)	0,64	0,34
självkänsla (10)	0,48	-0,07
familjefunktion (10)	0,36	0,17
förbättring av yrkesskicklighet (10)	0,22	0,02
rädsla för brottspåföljd (10)	0,12	-0,15

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

Minimitröskelnivåer och »kaskadeffekter«⁷

Om programteori var väl utvecklad för en omfattande intervention skulle PTE-data kunna användas i metaanalys för att ge information till beslutsfattare. Till exempel skulle mellanliggande variabler och resultatvariabler kunna tillåta uppskattningar av minimitröskelnivåer, dvs. den förbättring som skulle krävas i den mellanliggande variabeln för att ge förbättrade resultat. Detta skulle kunna vara till hjälp för program där de mellanliggande variablerna mäts vid något tillfälle före resultaten. Ett misslyckande med att åstadkomma en mellanliggande effekt skulle kunna tjäna som en varningssignal för beslutsfattare om att programmet är på väg mot dåliga resultat (Weiss, 1997).

Sådana data skulle också kunna skildra »kaskadeffekter«, eller sannolikheten för att det för varje påföljande mellanliggande variabel eller länk i en programteorikedja troligen kommer att redovisas mindre effekter. I till exempel kunskap-attityd-beteende (KAB)-modellerna (Lipsey, 1997) visar program som redovisar stor effekt på kunskap vanligen mycket mindre effekter på attitydmätningarna och ännu mindre på beteenderesultaten. Data om dessa kaskadeffekter skulle kunna användas för att tala om för beslutsfattare att ett program behöver nya verktyg för att kunna åstadkomma större effekter på kunskaper eller attityder för att mer effektivt kunna förändra beteendet i ett senare skede. Eller också behö-

ver kanske modellen revideras för att få med andra mellanliggande variabler mellan länkarna attityd och beteende.

Mervärdet av PTE

Metaanalys skulle kunna ge en metod för uppskattning av det mervärde som PTE ger. Mervärde betyder ofta någonting som kan mätas matematiskt, men här syftar det på huruvida PTE ger någon fördel förutom att tillhandahålla andra sätt att gripa sig an utvärdering eller inte. Trots att fördelarna med PTE länge har antytts har de inte blivit empiriskt påvisade. Ett sätt att pröva mervärdet är genom metaanalys. Lipton (1995) och hans kollegor genomför till exempel en metaanalys av utvärderingar av kriminalvårdsprogram som redovisats sedan 1968. Deras metaanalys kommer sannolikt att innehålla över 1 000 utvärderingar, varav vissa använder andra metoder såsom PTE.

Deras data skulle kunna användas för att jämföra PTE med dessa andra utvärderingsmetoder. Utvärderingarna kanske inte är så lätta att kategorisera, men skulle kunna graderas längs ett kontinuum för hur väl utvecklad den teori är som används för att styra utvärderingen (Lipsey, 1988). Graderingarna skulle kunna analyseras för att bestämma påverkan av teoriutvecklingen inom PTE på ett antal beroende variabler, inklusive effektstorlek, programframgång eller -misslyckande och så vidare. Ett mindre antal av studierna skulle kunna undersökas för att bestämma hur de användes i de påföljande besluten. Även om framtagna data bara är ungefärliga skulle de kunna ge vägledning beträffande

7 Mark Lipsey föreslog termen »kaskadeffekt« vid personlig kontakt i april 1998.

de reella fördelarna som PTE ger avseende processresultat eller utvärderingar enligt ettstegsmodellen.

Hinder och begränsningar

Det är inte många författare som skulle ha invändningar mot att ta med mellanliggande variabler vid utformningen av en utvärdering (t.ex. Cook, 2000). Men likaväl som det finns hinder mot att genomföra en enstaka PTE finns det vägspärrar uppsatta mot användningen av PTE i metaanalys. Bland dessa finns:

Det låga antalet PTE. Det största hindret är det låga antalet tillgängliga PTE som redovisats i facklitteraturen. Vårt eget sökande efter bra exempel på PTE var svårt (se Rogers et al., 2000). Till och med enkla PTE som kräver tvåstegsmodeller är svåra att hitta. Det är sällan som utvärderare tydligt och prospektivt redovisar en modell som kan prövas.

Betoningen på experiment och kvantitativa experiment. De flesta metaanalyser kräver som ett behörighetskriterium att de ursprungliga utvärderingarna innefattar en kontroll- eller jämförelsegrupp. Det här är ren kohandel eftersom man ökar den interna validiteten men utesluter potentiellt användbara studier som använder andra metoder för att utvärdera program.

Dålig redovisning. Granskare är allmänt missbelåtna med den dåliga redovisningskvaliteten i originalforskningshandlingar. Kombinationen av PTE och metaanalys skulle kräva att mer data samlades ihop, analyserades och redovisades av utvärderarna. Alla rekommenderar en förbättring

av redovisningskvaliteten men det har varit svårt att hitta lösningar.

Förenklade programteorier. Den här artikeln har inte tagit hänsyn till komplicerade modeller. De enkla PTE som diskuteras här är linjära och förutsätter en dominoeffekt – en förändring i en variabel kommer att resultera i en påföljande förändring i nästa mätta variabel. Som Rogers (2000) noterar fungerar kanske inte världen på det sätt som dessa modeller antyder. Modeller kan vara utförliga även i linjära teorier. Weiss (1997) förtecknar 17 länkar i sitt yrkesutbildningsexempel. Som hon noterar om utvärderare som genomför originalstudier (2000), kan metaanalytiker också tvingas bestämma vilka länkar i vilken teori de skall koda och undersöka vid sina genomgångar.

En rekommendation att främja PTE

Som nämndes tidigare är det huvudsakliga hindret för den här metoden bristen på PTE. Sherman och hans kollegor (1997) föreslår en metod för att öka antalet högkvalitativa utvärderingar. I deras genomgång av brottsförebyggande studier för den amerikanska kongressen undersökte de också de utvärderingskrav som ställs av den federala regeringen och delstatsregeringarna när de finansierar program som är relaterade till rättskipning i brottmål. Trots att det som en förutsättning för finansieringen vanligtvis krävs att alla anslagsmottagare skall genomföra en utvärdering fann Sherman och hans kollegor att bara ett fåtal finns redovisade. Ett problem är att det som

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

kallas för utvärdering ibland bara är input-data eller information om betjänade kunder. I det fåtal resultat- eller påverkansstudier som genomförs ingår sällan kontroll- eller jämförelsegrupper. Kontentan är att mycket litet är känt om vad det är som fungerar inom brottsförebyggandet (Sherman et al., 1997). En bidragande orsak är den i allmänhet otillräckliga mängden pengar för utvärdering som finns med i programbudgeten. De föreslår en annan metod: I stället för att kräva utvärdering av alla finansierade program bör det administrerande organet (vanligen på federal nivå eller delstatsnivå) slå samman utvärderingsmedlen för att stödja ett mindre antal utvärderingar av hög kvalitet på bara några få platser (t.ex. Sherman et al., 1997). En sådan metod skulle kunna bidra till att främja en ökning av antalet mycket noggrant styrda PTE.

Ett exempel på hur den här strategin skulle kunna fungera i praktiken kommer från ett program finansierat av federala medel.⁸ Titel V-anlagen för lokala brottsförebyggande program från Office of Juvenile Justice and Delinquency Prevention (OJJDP) stödde 477 särskilda interventioner i USA under räkenskapsåren 1994-1997 (OJJDP, 1998). I stället för att kräva en utvärdering i alla de 477 fallen (budgeten uppgick vanligtvis till 10 000 dollar per anslagsställe) skulle de 4,7 miljoner dollar

som avsatts till utvärdering kunna användas till finansiering av PTE i 20 av fallen. Varje PTE skulle kunna innefatta randomiserade kontrollerade studier eller någorlunda meningsfulla jämförelsegrupper. Varje utvärdering skulle kunna få 200 000 dollar per anslagsställe (till en total kostnad av 4 miljoner dollar). De andra medlen (700 000 dollar) skulle kunna användas för att bygga in en begränsad samling av data för kontroll av de ställen som inte har PTE. Om man använde sig av det här tillvägagångssättet skulle en metaanalys av de 20 PTE:na kunna utföras på en rimligt kort tid. En systematiskt och mycket noggrant utförd granskning av 20 PTE skulle med all säkerhet kunna ge mycket bättre information än de 477 lågkvalitativa och utspridda utvärderingar som annars ofrånkomligen skulle komma att redovisas.

8 De statliga byråerna förmedlar vanligtvis anslagen i olika «finansieringsriktningar». USA:s justitiedepartement har t.ex. många finansieringsriktningar (t.ex. lagen om kvinnovåld). Anslagen fördelas blockvis till delstaterna för de olika riktningarna och sedan fördelar delstaterna dem i sin tur som underanslag (Sherman et al., 1997).

Referenser

- Bickman, Leonard, Editor (1987) *Using Program Theory in Evaluation*. San Francisco, CA: Jossey-Bass.
- Chandler, MJ (1973) Egocentrism and antisocial behavior: The assessment and training of social perspective-taking skills. *Developmental Psychology* 9, 3, 326-332.
- Chen, Huey T. and Peter Rossi. Editors (1992) *Using Theory to Improve Program and Policy Evaluations*. New York: Greenwood.
- Cook, Thomas D. (2000) The false choice between theory-based evaluation and experimentation *New Directions in Evaluation* 87, 27-34.
- Cordray, David S. (1992) Theory-driven meta-analysis: Practices and prospects In Chen, H.T., Rossi, P. ed.: *Using Theory to Improve Program and Policy Evaluations*. New York: Greenwood.
- Davies, P., A. Petrosino and I. Chalmers (1999) *Proceedings of the International Meeting on Systematic Reviews of the Effects of Social and Educational Interventions*. July 15-16, London: University College-London, School of Public Policy.
- Donaldson, Stuart I. (2001) Mediator and moderator analysis in program development In Sussman, S, Editor: *Handbook of Program Development in Health Behavior Research and Practice*. Newbury Park, CA: Sage.
- Greater Egypt Regional Planning & Development Commission (1979) *Menard Correctional Center. Juvenile tours impact study*. Carbondale, IL: Greater Egypt Regional Planning & Development Commission.
- Hunt, Morton (1997) *The Story of Meta-Analysis*. New York: Russell Sage Foundation.
- Laws, Richard, Editor (1989) *Relapse Prevention with Sex Offenders*. New York: Guilford.
- Lipsey, MW (1997) What can you build with thousands of bricks? Musings on the cumulation of knowledge in program evaluation. *New Directions in Evaluation* 76,7-24.
- Lipsey, M. W. (1990) Design sensitivity: Statistical power for experimental research. Newbury Park, CA: Sage.
- Lipsey, MW and Wilson, DB (1993) The efficacy of psychological, educational and behavioral treatment. Confirmation from meta-analysis. *American Psychologist* 48, 12, 1181-1209.
- Lipsey, MW and Pollard JA (1989) Driving toward theory in program evaluation: More models to choose from. *Evaluation and Program Planning* 12,317-329
- Lipsey, MW (1988) Practice and malpractice in evaluation research. *Evaluation Practice* 9, 4,5-24.
- Lipton, DS (1995) CDATE: Updating the Effectiveness of Correctional Treatment 25 years later. *Journal of Offender Rehabilitation*, 22,(1/2),1-20.
- Petrosino, A (2000) Answering the why question in evaluation: The causal-model approach. *Canadian Journal of Program Evaluation* 12, 1, 1-25.
- Office of Juvenile Justice and Delinquency Prevention (1998) *1998 Report to Congress Title V Incentive Grants for Local Delinquency Prevention Programs*. Washington, DC: OJJDP.
- Rogers, Patricia (2000) Causal models in program theory evaluation. *New Directions in Evaluation* 87, 47-56.
- Rogers, Patricia, Anthony Petrosino, Tim Hacsí and Tracy Huebner (2000) Program theory evaluation: practice, promise, problems. *New Directions in Evaluation* 87, 5-14.
- Rosenthal, R, and Rubin, D (1982) A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology* 74,166-169.
- Sherman, Lawrence W., Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter and Shawn Bushway (1997) *Preventing Crime: What Works, What Doesn't, What's Promising. A Report to the United States Congress*. College Park, MD: University of Maryland, Department of Criminology and Criminal Justice.
- Stufflebeam, Daniel, Editor (2001) *Evaluation*

Anthony Petrosino: Om och varför? Den potentiella nyttan av att inkludera....

- Models. *New Directions for Evaluation* 89. San Francisco, CA: Jossey-Bass.
- Weiss, Carol H. (1972) *Evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
- Weiss, CH (1997) Theory-based evaluation: Past, present, and future. *New Directions for Evaluation* 76,41-55
- Weiss, Carol H. (2000) Which links in which theories shall we evaluate? *New Directions in Evaluation* 87, 35-46.
- Wilson, DB, Gallagher, CA and MacKenzie, DL (2000) A meta-analysis of corrections-based education, vocation and work programs for adult offenders. *Journal of Research in Crime & Delinquency* 37, 4, 347-368.

Summary

Whether and why? The potential benefits of including program theory evaluations in meta-analysis

Over the past thirty years, two different methodologies have been promoted to resolve persistent challenges to those attempting to ascertain the effectiveness of social programs and policies. Many argue that it is not enough to know whether a program works or not, but why it succeeds or fails. What are the critical mechanisms or mediating influences involved? Over the past three decades, scientists have proposed theory-based or theory-driven (i.e. program theory) evaluation as a method for understanding more about why a program works or fails. Another challenge is how to handle the results from separate but similar studies. Meta-analysis, or the quantitative analysis of prior research studies, has been promoted as one method for resolving the difficulty of accumulating the results from prior evaluation studies. In this paper, the author proposes merging these two methodologies to provide even more benefit.

The author first describes simple one-step models that examine the effects of a program on an ultimate or distal outcome only. The paper then contains an illustra-

tion, using hypothetical data, on how such one-step models can affect meta-analysis. In response, program theory evaluations provide further insight on the critical mechanisms that mediate the relationship between the program and the ultimate outcome. The author uses both hypothetical and real world data to show how this information can be used to illuminate our understanding of social policy and program impact. For program theory evaluations to be used in meta-analysis, they must meet a set of minimum requirements, including the measurement of the impact of the program on at least one mediating variable and one outcome variable. Other benefits suggested by this merger include informing larger social science theories, identifying minimum threshold levels and cascading effects, and specifying the value-added by theory-based evaluation. Limitations are also discussed. The author concludes with one recommendation for increasing the number of high-quality program theory evaluations that can later be analyzed in research synthesis.