

Avaluació *ex ante*: la importància d'un estudi base

Sebastià Sarasa Urdiola

PID_00197903



Els textos i imatges publicats en aquesta obra estan subjectes –llevat que s'indiqui el contrari– a una llicència de Reconeixement-NoComercial-SenseObraDerivada (BY-NC-ND) v.3.0 Espanya de Creative Commons. Podeu copiar-los, distribuir-los i transmetre'ls públicament sempre que en citeu l'autor i la font (FUOC. Fundació per a la Universitat Oberta de Catalunya), no en feu un ús comercial i no en feu obra derivada. La llicència completa es pot consultar a <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.ca>

Índex

Introducció	5
1. Estudi base: l'anàlisi del problema	7
1.1. Estimar l'abast i la rellevància del problema	7
1.2. Anàlisi de les causes del problema	7
1.3. Valorar les evidències empíriques disponibles sobre l'eficàcia que tenen programes similars al que es vol posar en marxa	8
2. La revisió sistemàtica dels estudis publicats	10
2.1. Definir de manera concisa la pregunta	10
2.2. Cercar el màxim nombre possible d'estudis que donin resposta a les preguntes formulades	11
2.3. Filtrar i seleccionar els estudis que siguin rellevants	12
2.4. Valorar la qualitat dels estudis seleccionats	13
2.5. Analitzar i comparar els resultats que ha obtingut cada avaluació	13
3. La metanàlisi	15
3.1. Resultats mesurats com a diferències de mitjanes entre el grup participant i el grup de comparació o de control	16
4. Lectures complementàries	21

Introducció

En la bibliografia sobre avaluació de programes es fa una distinció entre l'estimació dels resultats esperats abans d'invertir els recursos en un programa, i l'avaluació que es fa en un moment donat de la vida del programa per valorar quins han estat els resultats reals de la intervenció. La primera avaluació es coneix com a *avaluació ex ante*, i a la segona com a *avaluació ex post* (Rossi i altres, 2004). L'avaluació de l'impacte que ha tingut un programa és l'objecte d'aquesta assignatura, per tant, l'assignatura es focalitza en l'avaluació *ex post*, però l'estudiant ha de ser conscient de com de condicionada està una avaluació *ex post* fiable d'una avaluació *ex ante* rigorosa. És per aquesta raó que esmerçarem uns mots a entendre les bases de l'avaluació *ex ante*.

L'avaluació prèvia a un programa ajuda a identificar i avaluar les necessitats de la població que ha de ser atesa. Ajuda també a garantir que el programa tindrà uns objectius clars, que aquests objectius seran assolibles, que els recursos emprats seran eficaços, i que **una avaluació fiable, *ex post*, de l'impacte del programa és possible**. L'avaluació *ex ante* és la base sobre la qual se sustenta el bon disseny d'un programa, però a nosaltres ens interessa la part de l'avaluació *ex ante* que condicionarà l'avaluació *ex post*, la que revisa la validesa de les teories i de les evidències empíriques sobre les quals se sustentaran les hipòtesis predictives dels efectes esperats del programa que haurem d'avaluar *a posteriori*. Els objectius del programa són els resultats que esperem assolir, però aquesta esperança és versemblant quan els resultats esperats estan fonamentats en hipòtesis avalades per teories i evidències empíriques d'estudis previs, com també per les avaluacions que s'han fet d'altres programes similars al que volem desenvolupar.

És per això que, abans del disseny d'un programa, és recomanable executar tres tasques bàsiques: una anàlisi prèvia del problema, una avaluació de les necessitats i una estimació dels recursos que es requereixen. D'aquestes tres tasques, la primera, l'anàlisi del problema, és especialment rellevant per a fer possible una avaluació fiable *ex post*, i és en la que focalitzarem l'exposició d'aquesta secció de l'assignatura.

1. Estudi base: l'anàlisi del problema

Analitzar un problema, amb ànim de trobar una manera d'intervenir-hi, per a fer prevenció o per a atendre les necessitats dels afectats, demana abordar tres qüestions cabdals: estimar la rellevància del problema, analitzar-ne les causes i valorar les evidències empíriques d'estudis previs que puguin donar suport a la metodologia d'intervenció.

1.1. Estimar l'abast i la rellevància del problema

Un programa està més legitimat davant l'opinió pública i els possibles finançadors com més capaços siguem de demostrar que el problema que volem resoldre és rellevant per al benestar social de la col·lectivitat o d'un grup social específic. La rellevància d'un problema ens ve determinada per la quantitat de persones afectades, per la intensitat i caràcter irreversible del mal que provoca, i pels costos d'oportunitat que comporta.

Si prenem la pobresa infantil a Catalunya l'any 2011 com a exemple de problema social, la seva rellevància queda determinada per la seva extensió (aproximadament un de cada quatre menors de setze anys viu en una llar amb una renda disponible per sota del llindar de pobresa, una de les taxes més elevades de la Unió Europea). És sabut que el desenvolupament dels infants en ambients de pobresa té conseqüències desastroses, i irreversibles en molts casos, en la salut, formació, capacitat de relacionar-se socialment i, en suma, en les oportunitats vitals. Endemés, sabem que una nació on la pobresa infantil és gran, té més dificultats per a desenvolupar una força de treball que faci l'economia del país competitiva en el mercat internacional, perd capacitat de generar riquesa col·lectiva, i té costos associats de més despesa social i manca de cohesió social.

En conseqüència, mostrar la rellevància del problema que un programa vol prevenir ens obliga a contestar preguntes del tipus:

- Quina és la població afectada directament i indirecta per les conseqüències del problema?
- Quins són els costos socials que ha de suportar la comunitat a causa del problema?

1.2. Anàlisi de les causes del problema

En el mòdul "Els reptes principals d'una avaluació d'impacte", hi ha una reflexió més detallada sobre què es considera una explicació causal a partir de la teoria i les evidències empíriques disponibles sobre un problema; ara, però, convé que fem una descripció de com hem de fer front de manera analítica a un problema que volem pal·liar o prevenir.

Si continuem amb l'exemple anterior de la pobresa infantil, un cop hem estimat l'abast i conseqüències que té la pobresa al nostre país, ens hem de preguntar sobre quins són els agents implicats en aquesta situació, a més, és clar, dels mateixos infants.

Assumim, per fer l'exemple menys complicat, un concepte de *pobresa* que es limita a obtenir una renda monetària insuficient per a cobrir despeses bàsiques. Des d'aquest enfocament, els agents més destacats són, en primer lloc, els pares i mares que o no estan

ocupats en el mercat de treball o que, estant-ho, obtenen ingressos molt baixos. D'aquests pares i mares, uns estan en aquesta situació perquè les seves capacitats són minses i tenen poques oportunitats laborals (limitacions de capacitats que poden tenir l'origen en problemes de salut o de formació deficient). D'altres, però, poden estar en situació de pobresa perquè una crisi econòmica ha destruït part del teixit productiu i ha deixat sense ocupació un volum de treballadors assalariats, treballadors autònoms i petits empresaris que mai abans no havien tingut problemes econòmics. D'altra banda, hi ha nens pobres, no perquè els seus pares no estiguin ocupats, sinó perquè tenint una feina reben uns salaris molt baixos.

Entre els agents implicats, a més dels pares que cobren salaris baixos o no troben feina, hi ha les administracions públiques dels diferents nivells –estatal, autonòmica i local–, cadascuna amb competències diferents tant pel que fa a la regulació del mercat de treball i de l'activitat econòmica, com a la capacitat per a oferir ajudes econòmiques a les llars pobres i subsidis per a consumir serveis escolars, preescolars i de lleure que ajudin a pal·liar les necessitats dels infants pobres. A més, hem de considerar el paper de les empreses i les condicions de l'economia del país que afavoreixen determinades ocupacions i en penalitzen altres. Veiem, doncs, que una tasca a fer en l'anàlisi del problema és determinar les diferents tipologies de pobresa infantil, ja que la solució al problema serà diferent depenent de la causa de la pobresa.

Si hi ha un canvi en el cicle econòmic que augmenta la demanda de força de treball, molt probablement molts dels menors, fills de pares amb qualificacions intermèdies i elevades, sortiran de la pobresa perquè els seus pares tornaran a estar ocupats. Més difícil serà per als treballadors amb baixa qualificació o amb discapacitats mantenir una relació laboral estable (persones malaltes, dependents de substàncies addictives, amb dificultats de relació social, etc.), raó per la qual un programa de lluita contra la pobresa en aquest grup poblacional no pot tenir gaire confiança en l'evolució del mercat laboral. D'altra banda, malgrat que hi hagi una millora en el mercat laboral, continuaran havent-hi mares que no treballin si no es fa res per a modificar les condicions que abans, en el moment de bonança econòmica, ja les feia ser inactives com, per exemple, salaris baixos i costos elevats dels serveis substitutius de la seva activitat a la llar.

Com veiem, una bona anàlisi del problema demana un treball exhaustiu en el qual hem de considerar dues qüestions:

- Quins són els agents implicats en el problema?
- Quines són les causes del problema?, (teories que disputen l'explicació de la lògica situacional que condiciona la interacció dels agents als seus interessos, creences i valors, etc.).

Les dues preguntes, però especialment la segona, ens demanen una elaboració completa d'allò que podríem anomenar *l'estat de la qüestió*, és a dir, una revisió d'allò publicat en les revistes científiques més prestigioses en el camp que ens interessa, per a identificar quines són les teories principals i les evidències empíriques que els donen suport. Aquest estat de la qüestió és el que ajudarà a perfilar quines són les causes més rellevants del problema que ens interessa afrontar amb el programa i, en conseqüència, ens servirà per a fixar els objectius que raonablement pot assolir el programa i que hauran de ser la referència per a avaluar posteriorment l'impacte.

1.3. Valorar les evidències empíriques disponibles sobre l'eficàcia que tenen programes similars al que es vol posar en marxa

Un cop revisat el marc teòric que explica les causes del problema, i seleccionades les teories que semblen més raonables, fiables i contrastades amb prove empíriques, queda la tasca de revisar la bibliografia científica especialitzada en

avaluació de projectes per a cercar les avaluacions fetes de programes similars al que volem posar en marxa. És el moment de formular-nos preguntes del tipus següent:

- Hi ha programes que hagin intervingut en problemes similars?
- Han estat avaluats?
- Quin és el grau de rigor d'aquestes avaluacions?
- Quines conclusions fiables en podem treure?
- Quina previsió de resultats estimem del nostre programa a partir dels resultats dels altres?

La recerca en aquest camp consta de dues fases: primer, una revisió sistemàtica dels estudis fets i, després, si és possible, una metanàlisi dels resultats.

2. La revisió sistemàtica dels estudis publicats

A l'hora de definir un programa hem de recórrer a l'experiència acumulada per la comunitat científica que ha avaluat altres programes similars prèviament. La idea de la revisió sistemàtica consisteix a acceptar que el coneixement avança de manera acumulativa i que poca contribució farem a la millora del benestar amb un programa si no tenim en consideració les experiències i els coneixements que han acumulat altres abans que nosaltres.

La qüestió és com s'accedeix a la informació, esbrinar si mai han estat avaluats programes similars i on podem trobar-ne els resultats, i com podem esbrinar si la informació és fiable. Quan fem una revisió sistemàtica és prudent que seguim els passos que detallem a continuació:

- Definir de manera concisa la pregunta.
- Cercar el màxim nombre possible d'estudis que donin resposta a les preguntes formulades.
- Filtrar i seleccionar els estudis que siguin rellevants.
- Valorar la qualitat dels estudis seleccionats.
- Analitzar i comparar els resultats de cada estudi.

2.1. Definir de manera concisa la pregunta

Hem de definir la pregunta que ens interessa contestar relativa als resultats esperats del nostre programa. No té sentit començar la cerca de publicacions científiques si no tenim clar què és allò que estem cercant. Si la nostra pregunta, tornant a l'exemple de la pobresa infantil, és de tipus genèric com ara sota quines condicions la pobresa infantil és més baixa, buscarem una informació ben diferent de si la pregunta formulada és quant redueix la pobresa infantil un programa de subsidis públics en escoles bressol. La pregunta ha d'incorporar subpreguntes del tipus següent:

- Quina població ha de ser objecte de l'estudi? Un grup d'edat determinat? Una població d'una classe social concreta? Una població caracteritzada per una necessitat específica?
- Quina mena de resultats ens interessin? Si, per exemple, estem interessats a valorar l'eficàcia de programes que treuen de la pobresa la infància, haurérem de rebutjar altres estudis que no tinguin en compte aquest resultat, malgrat que siguin programes que millorin el benestar dels menors. Cal incloure, però, els resultats adversos que hagin pogut produir les intervencions similars a la que volem fer. No és estrany que una intervenció tingui efectes imprevistos, no buscats, que fins i tot siguin contraproductius. En

Pàgina web

Una font d'informació rigorosa en què l'estudiant pot trobar un depòsit de revisions sistemàtiques en diferents àmbits específics de les ciències socials és el centre de recursos de The Campbell Collaboration.

Vegeu també

Vegeu la diferència entre aquests mètodes en el mòdul "Metodologies d'avaluació dels resultats de programes".

aquests casos, hem d'estar amatents de recollir els estudis en què els efectes adversos són documentats.

- Quina mena de metodologies acceptem? Volem només estudis experimentals amb un grup experimental i un grup de control, o volem mètodes quasiexperimentals?

2.2. Cercar el màxim nombre possible d'estudis que donin resposta a les preguntes formulades

Una manera que tenim de discriminar les nostres lectures de treballs científics és verificar si ens ofereixen un estat de la qüestió sobre tot allò que ja és sabut i quina és l'aportació del treball que llegim a aquest coneixement a partir de les dades empíriques que maneja.

La millor manera d'evitar caure en la trampa de donar per bons uns resultats que no s'ajusten a la realitat, perquè han estat fruit de la casualitat atzarosa, és consultar el màxim nombre possible d'estudis. Serà difícil que una mateixa casualitat es repeteixi de manera recurrent en la majoria dels estudis. Però hem de tenir cura que la nostra cerca no estigui esbiaixada per alguna raó, per exemple, perquè rebutgem revisar les publicacions d'una disciplina que, per ser lluny de la nostra formació, obviem. Penseu els lectors, per exemple, en investigadors formats en una disciplina (l'economia, la psicologia o la sociologia) que a l'hora de cercar estudis previs només consulten revistes científiques de la seva pròpia disciplina, menyspreant implícitament les perspectives i informació que aporten altres ciències socials. O, desgraciadament força freqüent al nostre país, investigadors que, per manca de domini de la llengua anglesa, limiten la seva revisió a les referències publicades en castellà i català; abandonant la informació més actual i àmplia publicada en les revistes en què la comunitat científica internacional debat els seus estudis.

Una bona estratègia per a accedir als estudis científics és consultar alguna de les bases electròniques que ordenen les revistes de cada disciplina a partir de la freqüència en què són citats els articles que publiquen.

Una bona pràctica consisteix a baixar primer d'Internet només els resums dels articles i llibres (*abstracts*), ja que ens podem trobar amb un volum de treballs considerable i costós econòmicament si hem d'adquirir i pagar els articles. Aquests resums ens han de subministrar prou informació per a valorar si paga la pena baixar l'article sencer o no.

Dos consells a seguir:

- Primer, cercar si hi ha revisions sistemàtiques ja publicades. Si n'hi ha, ens estalviaran una bona part de la feina, ja que ens quedarà com a llacuna

Journal Citation Report Social Science Edition

En aquesta base de dades podem veure quin ha estat cada any l'ordre de les revistes més citades segons la temàtica que ens interessen. Un cop seleccionades les revistes més rellevants en el rànquing, podem començar la cerca dels articles que ofereixen els treballs que cerquem.

principal, les publicacions possibles entre el moment actual i el moment en què es va fer la darrera revisió sistemàtica.

- Segon, resseguir les referències bibliogràfiques fetes en els articles que hem seleccionat sobre altres treballs que estan relacionats amb les preguntes que ens interessin.

Aquest procediment opera com una bola de neu que va captant noves referències fins que s'esgoten, i té l'avantatge que ens permet identificar quins han estat els estudis més rellevants pel nombre de vegades que han estat citats per la comunitat científica.

Una advertència, però: en general, hi ha un risc de biaix a causa del que podríem anomenar *l'efecte publicació*. Un efecte degut al fet que la informació publicada pot estar esbiaixada per diverses raons com ara interessos de certs grups a defensar alguns resultats i amagar els fracassos, incompetència dels seleccionadors d'allò que és publicable en algunes revistes, o simplement perquè hi ha una tendència a publicar només els resultats que han tingut èxit i a deixar a l'ombra les avaluacions que no confirmen els resultats esperats. Per això és una bona idea cercar documents de treball no publicats en revistes i incorporar-los, sempre que la causa de no haver-los publicat no sigui la manca de rigor. Una manera d'accedir a documents de treball és fent consultes a experts en el tema que ens indiquin els centres de recerca on podem trobar-los.

2.3. Filtrar i seleccionar els estudis que siguin rellevants

Un cop s'ha fet la selecció prèvia dels articles que tracten el tema que ens interessa, es procedeix a una tasca que és molt delicada: la selecció final que recollirà els articles a analitzar i que exclourà els altres.

Les raons per a seleccionar uns treballs i excloure'n d'altres han d'estar relacionades amb una revisió més estricta sobre si els articles donen resposta a les preguntes que ens havíem formulat sobre població en què s'ha intervingut, resultats mesurats i metodologies d'avaluació. Quan un mateix estudi ha estat publicat en diverses revistes, o col·lecció de papers de treball (*working papers*), només hem de triar la versió que aportí més informació sobre les preguntes que ens interessin i rebutjar les altres versions, ja que ens interessa avaluar els resultats dels estudis fets, no els articles que s'hagin publicat.

En aquesta fase, els avaluadors *ex ante* han de ser molt curiosos i informar de manera documentada sobre les raons per les quals uns articles han estat rebutjats de l'anàlisi, de manera que els observadors externs puguin jutjar el grau en què la revisió sistemàtica no ha estat esbiaixada.

No obstant això, qui fa una revisió sistemàtica ha d'estar sempre amatent als efectes que les seves decisions sobre quins treballs s'accepten tenen en els resultats finals de la revisió. Per això és prudent contrastar la sensibilitat que tenen

les conclusions de la revisió sistemàtica a la inclusió o exclusió de determinats estudis. Si comprovem que l'exclusió d'un determinat tipus d'estudis altera les conclusions de manera significativa, haurem de ser curosos en l'argumentació en defensa de l'exclusió i, en cas de tenir dubtes, és millor oferir les dues conclusions i deixar per a una fase posterior de debat amb altres responsables del programa i experts externs quina opció triar.

2.4. Valorar la qualitat dels estudis seleccionats

La qualitat dels estudis s'ha de valorar per la validesa de les seves conclusions. No té gaire sentit esmerçar energies a comparar els resultats de programes que o bé estan esbiaixats per defectes metodològics (validesa interna) o bé s'han desenvolupat en unes condicions que no els fan generalitzables a tota la població, o a la població en la qual volem intervenir (validesa externa).

Vegeu també

Per a revisar els conceptes de *validesa interna* i *externa*, vegeu el mòdul "Els reptes principals d'una avaluació d'impacte".

2.5. Analitzar i comparar els resultats que ha obtingut cada avaluació

De cada estudi revisat convindrà indicar en una graella els resultats que ha aconseguit el programa que s'avalua. La manera de presentar els resultats dependrà de la metodologia d'estudi. Quan es tracta d'avaluacions seguint el mètode experimental, en què la similitud entre els individus del grup experimental i els del grup de control està garantida per l'assignació aleatòria a cada grup, n'hi ha prou de recollir els valors de la variable dependent (el resultat del programa) en tots dos grups de l'estudi, la qual cosa ens permetrà després comparar les diferències entre el grup experimental i el grup de control en tots els estudis seleccionats. Com es recullen els valors de la variable dependent estarà condicionat pel tipus de variable que sigui.

Vegeu també

Hem vist els mètodes experimental i quasiexperimental en el mòdul "Metodologies d'avaluació dels resultats de programes".

Hi ha programes en què el resultat es mesura amb **variables categòriques dicotòmiques** com a = ha tingut un resultat positiu, b = ha tingut un resultat negatiu, o amb més categories, com per exemple en un programa per a promoure l'activitat entre joves inactius: a = ha obtingut una ocupació; b = ha començat a estudiar de nou; c = continua sense estudiar ni treballar. En aquests casos, s'introduirà a la graella, en la línia de cada avaluació, el nombre d'individus que formen part del grup que participa en el programa i el nombre d'individus que formen part de grup de control. De cada grup, s'indicarà la proporció de casos que acaben en cada categoria.

Altres programes mesuren els resultats amb una **variable contínua**, com ara el nivell de renda de la família o les hores d'estudi setmanals. En aquests casos, la informació que s'ha d'enregistrar és la mitjana i la desviació estàndard que cada grup assoleix en l'estudi revisat.

Més difícil és trobar una recomanació universal sobre com s'han de recollir els resultats de les avaluacions quan la metodologia és quasiexperimental, ja que un mateix programa es pot haver avaluat seguint metodologies diferents amb

resultats no directament comparables. Fóra ideal que, en cas que el resultat del programa es mesurés amb variables categòriques, els resultats de l'avaluació es poguessin expressar en forma d'efectes parcials provocats per la variable *intervenció*, i, en cas que el resultat del programa es mesurés amb variables contínues, que els resultats de l'avaluació es poguessin expressar en forma d'efectes marginals. Tant l'efecte parcial d'una variable dicotòmica com l'efecte marginal d'una variable contínua es poden estimar amb les aplicacions informàtiques d'anàlisi estadística més comunes, però si l'estudiant té una formació estadística poc sofisticada, ha de saber que estem parlant del fet que el resultat d'una avaluació s'ha d'expressar en els termes següents: "controlats altres factors que puguin influir en el resultat, un augment d'una unitat en el tractament provoca un augment de X unitats en la variable *resultat*".

3. La metanàlisi

La fase final de la revisió sistemàtica enfocada a comparar els resultats obtinguts per les avaluacions de programes que han estat revisades pot anar un pas més enllà amb la tècnica de la metanàlisi, tècnica estadística de gran utilitat quan els programes avaluats són nombrosos i sobrepassen la nostra capacitat de valorar-los de manera intuïtiva. La metanàlisi constitueix una varietat més sofisticada de la comparació de resultats, però que no sempre és aplicable en tant que els estudis revisats potser no han estat publicats oferint el detall estadístic que necessitem per a fer una metanàlisi. Convé matisar que la metanàlisi és una tècnica aplicable sempre que volem estimar l'efecte mitjà d'una variable explicativa sobre una variable dependent; per tant, es tracta d'una tècnica de gran utilitat en l'avaluació *ex ante*, però que no és d'aplicació exclusiva en aquesta fase de la formació de programes. Com succeeix també amb la revisió sistemàtica en general, es tracta d'una tècnica que permet a un investigador obtenir de manera sintètica l'estat de qualsevol qüestió que sigui d'interès científic i que s'hagi investigat prèviament amb dades empíriques.

L'avantatge atribuït a la metanàlisi, respecte a la mera revisió sistemàtica, és que redueix bastant els errors deguts a la subjectivitat de qui fa la revisió sistemàtica de les avaluacions. La metanàlisi és una tècnica estadística que permet obtenir una síntesi dels resultats obtinguts per altres programes, a la vegada que obre la possibilitat de donar explicacions de per què els resultats són heterogenis entre programes. L'aportació principal de la metanàlisi a l'avaluació de programes és que ens permet valorar els resultats obtinguts per programes similars al que volem aplicar, programes que han estat desenvolupats en llocs diferents i amb poblacions diferents, la qual cosa ens permet contrastar fins a quin punt aquests resultats són generalitzables.

La metanàlisi consisteix a construir una base de dades que serà objecte de tractament estadístic, en què cada registre de la base recull dades referents als resultats obtinguts en cada avaluació.

La idea base de la metanàlisi és que, de la mateixa manera que podem fer estimacions sobre els valors mitjans de les característiques que té una mostra d'individus enregistrats en una base de dades, també podem fer estimacions dels valors mitjans, intervals de confiança i significació estadística dels resultats obtinguts per múltiples programes que persegueixen objectius similars.

La metanàlisi permet fer una síntesi de la direcció i la grandària de l'impacte que han tingut altres programes similars al que volem posar en marxa, i que han estat avaluats prèviament. També permet establir el grau de significació estadística que tenen els resultats obtinguts pels programes.

Les fórmules que s'han d'utilitzar depenen de com es mesuri la variable que indica els resultats dels programes, ja que aquests resultats es poden mesurar de diferents maneres, per exemple:

- diferències de mitjanes entre el grup participant i el de comparació;
- correlacions entre la variable dependent (resultat) i la variable independent (intervenció);
- diferències de casos d'èxit entre el grup participant i el grup de comparació;
- escales ordinals;
- funcions de supervivència, etc.

Atès el caràcter introductorï d'aquest text, no explicarem amb detall totes les fórmules que requereix cada tipus de variable, sinó que solament en descriurem breument una: la diferència entre mitjanes de variables contínues.

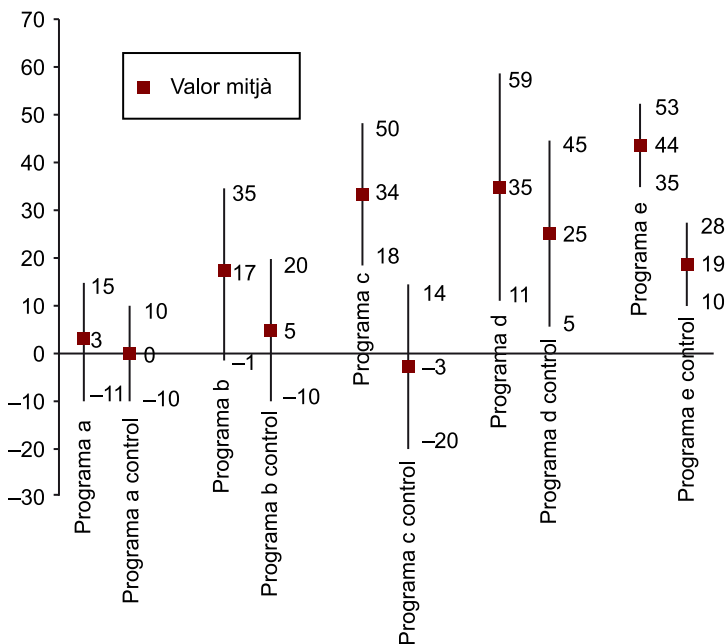
3.1. Resultats mesurats com a diferències de mitjanes entre el grup participant i el grup de comparació o de control

Hi ha moltes situacions en què el resultat esperat d'un programa el mesurem amb variables contínues del tipus mitjana d'hores, de dies o de qualsevol altra mesura temporal que els participants han dedicat a una activitat; mitjana de diners o de qualsevol altre recurs que els participants han obtingut; o mitjana d'unitats de consum que els participants han augmentat (per exemple en programes que volen millorar la dieta alimentària) o d'unitats de consum que deixen de consumir (com el tabac, o aliments contraindicats per a la seva salut). En aquests casos, s'han de comparar els valors mitjans dels participants amb els valors mitjans dels individus que han format part dels grups de comparació, o de control. D'aquesta comparació hem de concloure quina direcció i grandària és més probable que tinguin els efectes dels programes avaluats i la seva significació estadística.

Per exemple, en la figura 1 veiem els resultats hipotètics de cinc programes (anomenats *a*, *b*, *c*, *d*, *e*) orientats a augmentar el nombre d'hores trimestrals que els adolescents dediquen a l'esport mitjançant l'organització supervisada d'activitats de carrer. La gràfica ens mostra els valors mitjans dels increments en l'activitat esportiva, amb els seus intervals de confiança, dels participants en el programa i dels individus que formaven part dels grups de control. Podem veure que el valor mitjà dels participants en el programa *a* té un interval de confiança que va des de valors negatius a positius, la qual cosa ens fa pensar que no té un efecte clar en cap direcció. Les avaluacions dels altres programes mostren que els participants sí que han tingut un augment positiu del nombre d'hores mensuals dedicades a l'esport, però, només el programa *c* i el programa

e tenen valors mitjans dels participants amb uns intervals de confiança que no se superposen amb els intervals de confiança dels valors mitjans obtinguts pels adolescents dels grups de control. Això no ens permet assegurar que el comportament dels participants en els programes *b* i *d* hagi estat diferent del comportament dels grups de control. Per tant, i a primera vista, haurem de concloure que, de cinc programes que han estat avaluats, només dos ofereixen resultats positius.

Figura 1. Valors mitjans dels participants i dels grups de control (intervals de confiança del 95%)



Aquesta gràfica ens mostra els valors mitjans i els intervals de confiança corresponents als resultats obtinguts en cinc programes avaluats (*a, b, c, d, e*) pel grup de participants i el grup de control. Podem observar com, a cadascun dels programes *a, b* i *d*, els valors del grup de participants i el de control tenen intervals de confiança superposats, de manera que no podem assegurar que hi hagi diferències significatives entre un grup i un altre. En el programa *e*, sí que observem un resultat superior entre els participants, que es mou dins un interval situat entre 35 i 53, clarament per sobre del resultat del grup de control, que es mou entre 10 i 28. Passa el mateix amb el programa *c*.

Ara bé, aquesta inspecció gràfica és més consistent quan la basem en un càlcul de l'efecte mitjà aconseguit pels cinc programes avaluats. La taula 1 ens mostra els valors amb els quals s'ha dibuixat la figura 1. La mitjana dels cinc programes l'estimem de la manera següent:

1) Primer, per a evitar els problemes derivats del fet que cada programa hagi avaluat els seus resultats en unitats de mesura diferents, hem d'estandarditzar les unitats de mesura de cada programa per a convertir-les a una unitat de mesura comuna: unitats de desviació típica de la població. Això ho aconseguim convertint les mesures en unitats de desviació típica seguint la fórmula següent:

$$d = \frac{x_p - x_c}{dt}$$

en què

Distribució de resultats

Tingueu en consideració que, assumint que els resultats del programa tenen una distribució normal, el 95% dels valors de la distribució estaran compresos en l'interval format pel valor mitjà $-1,96 + 1,96$ desviacions típiques.

d = diferència de resultats tipificats

x_p = mitjana dels resultats dels participants en el programa

x_c = mitjana dels resultats dels membres del grup de comparació, o de control

dt = desviació típica del grup de comparació o de control.

És a dir, en el cas del programa a , la diferència estandarditzada de resultats entre el grup participant i el grup de comparació és la següent:

$$\frac{3-0}{5} = 0,60$$

2) Un cop calculada la diferència estandarditzada en cada programa, el pas següent és estimar la mitjana de tots els programes seguint la fórmula (1):

$$d_{mitjana} = \frac{\sum dt}{n}$$

en què

$d_{mitjana}$ = diferència mitjana de tots els programes

$\sum dt$ = suma de totes les diferències entre participants i grup de comparació de cada programa

n = nombre de programes avaluats

Aquesta fórmula, aplicada a les dades de la taula 1, ens dona el resultat següent:

$$\frac{0,60 + 1,60 + 4,35 + 1,00 + 8,33}{5} = 3,17$$

Ara bé, quan calculem l'interval de confiança del 95% per a aquest valor mitjà, ens dona que el valor es mou entre +4,5 i -4,5; és a dir que el valor mitjà oscil·la entre $3,17 - 4,5 = -1,33$, i $3,17 + 4,5 = 7,67$ incloent-hi valors negatius i, per tant, no podem descartar que l'efecte dels programes sigui insignificant.

Valor mitjà

Per a estimar l'interval de confiança de la mitjana dels efectes (dt) cal estimar-ne primer la variància, que serà:

$$Var(dt) = \frac{1}{\sum w_i}$$

en què w_i és la inversa de la variància $\frac{1}{v_i}$ i v_i és el quadrat de la desviació típica de la població sobre la qual es compara cada programa. En el cas de l'exemple que estem veient,

el programa *a* té una variància v de $S^2 = 25$, i una inversa de la variància $w = \frac{1}{25} = 0,04$. Per tant, calculant la inversa de la variància de cada programa, la variància de (*dt*) és:

$$\text{Var}(dt) = \frac{1}{0,04 + 0,01 + 0,02 + 0,01 + 0,11} = 5,26$$

La desviació típica de *dt* serà $\sqrt{5,26} = 2,29$, i l'interval de confiança es mourà entre $+1,96$ i $-1,96$ desviacions típiques; per tant, entre $+/-1,96 * 2,29 = +/-4,5$.

En suma, els resultats d'aquesta metanàlisi ens informen que els programes d'activitat supervisada al carrer orientats a augmentar la pràctica esportiva dels adolescents, i que segueixen un mètode similar al dels cinc programes avaluats, obtenen, de mitjana, un augment en el nombre d'hores trimestrals equivalent a 3,17 vegades la desviació estàndard de la mitjana de la població sobre la qual es faci el programa. Però aquest augment és poc fiable, en tant que pot oscil·lar entre $-1,33$ i $7,67$, és a dir, incloent-hi la possibilitat que no tingui cap efecte positiu sobre el comportament dels adolescents.

Mitjana

La mitjana calculada d'aquesta manera dóna la mateixa importància a tots els programes avaluats. El lector ha de considerar que hi ha procediments per a ponderar el pes de cada programa, per exemple, en funció de la variància de la mitjana, que està condicionada per la grandària de la mostra utilitzada en el programa, de manera que els programes amb més participants i menys variància estadística tenen més pes, en tant que les seves estimacions són més fiables i precises. La fórmula següent substituiria la fórmula (1), que hem utilitzat per a estimar l'efecte mitjà dels programes

$$d_{\text{mitjana ponderada}} = \frac{\sum w_i \times dt}{\sum w_i}$$

en què w_i és la inversa de la variància de cada programa $w_i = \frac{1}{v_i}$, i la variància v_i de cada programa és el quadrat de la desviació típica de cada programa que es mostra en la taula 1.

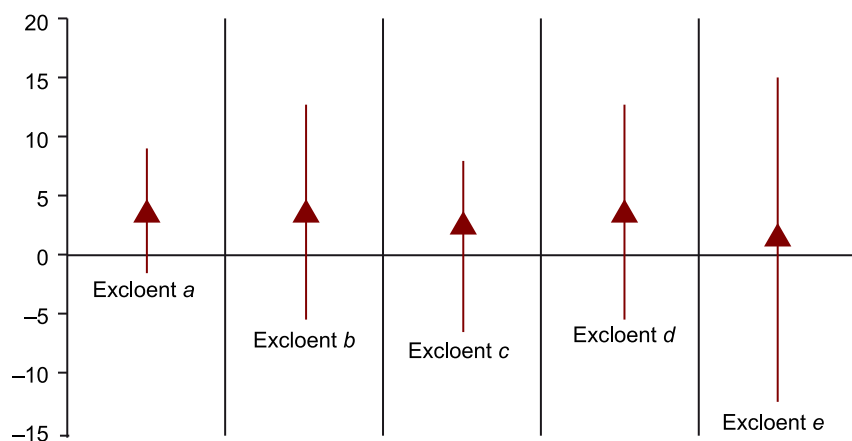
Taula 1. Resultats dels programes avaluats

	Interval de confiança 95%		Valor mitjà	Desviació típica	Diferència estandarditzada
Programa a	15	-11	3	6	0,60
Programa a control	10	-10	0	5	-
Programa b	35	-1	17	9	1,6
Programa b control	20	-10	5	7,5	-
Programa c	50	18	34	8	4,35
Programa c control	14	-20	-3	8,5	-
Programa d	59	11	35	12	1,00
Programa d control	45	5	25	10	-
Programa e	53	35	44	3	8,33
Programa e control	28	10	19	3	-

Estimar l'efecte mitjà d'un grapat de programes és una informació interessant per a qui ha de decidir si posar en marxa un programa similar, però els valors mitjans poden amagar realitats molt diferents. En el nostre exemple, hem vist que alguns programes tenen un efecte positiu clar, mentre que d'altres no. La qüestió, en una situació com aquesta, és esbrinar si aquestes diferències de

resultats són aleatòries o responen a alguna característica dels programes que en condicionen l'eficàcia. Per respondre a preguntes d'aquest tipus la metanàlisi ofereix algunes alternatives per a mesurar l'homogeneïtat dels programes avaluats. Entre aquestes, la més senzilla, quan el nombre de programes no és gaire gran, és l'anàlisi de sensibilitat. L'anàlisi de sensibilitat consisteix a repetir l'estimació de l'efecte mitjà tantes vegades com programes incloem en la metanàlisi, però deixant sempre un programa diferent fora de l'estimació. D'aquesta manera podem observar si l'efecte mitjà és alterat de manera significativa quan sostraiem un programa, o grup de programes, indicant-nos que no tots els programes són iguals. En la figura 2 podem veure l'anàlisi de sensibilitat aplicat als cinc programes hipotètics que hem avaluat. La gràfica ens mostra el valor de l'efecte mitjà amb el seu interval de confiança cada vegada que excloem un dels programes de l'estimació de l'efecte mitjà. El resultat ens mostra que els programes són força homogenis, ja que la sostracció de qual-sevol no altera gaire la significació de l'efecte mitjà que sempre es mou entre valors positius i negatius.

Figura 2. Anàlisi de sensibilitat



La gràfica ens mostra el valor mitjà dels resultats obtinguts per tots els programes menys un, que és exclòs del càlcul de la mitjana.

D'aquesta manera podem comprovar si l'exclusió del programa *a*, en la primera columna, modifica la mitjana de manera que deixi de ser no significativa, la qual cosa no succeeix, i tampoc quan excloem, un per un, els altres programes.

Arribem, doncs, a la conclusió que aquests programes metanalitzats no tenen resultats estadísticament significatius, i que aquesta manca de resultats és comuna a tots els programes.

Quan el nombre de programes que s'han d'avaluar és tan gran que fa feixuga l'anàlisi de sensibilitat, la metaregressió és una bona alternativa. La metaregressió és una regressió estadística en què els casos no són individus, sinó programes, la variable dependent és l'efecte dels programes i les variables independents són les característiques dels programes que teòricament poden ser la causa dels resultats heterogenis. Aquesta tècnica permet a l'avaluador de programes estimar quins factors poden explicar les diferències que hi ha entre programes pel que fa als resultats que s'han aconseguit. La metaregressió, però, està limitada en el nombre de variables independents que podem especificar en el model, ja que el nombre de programes sempre és molt limitat, i el nombre de variables independents que podem especificar en un model està condicionat pel nombre de casos que comparem.

4. Lectures complementàries

Asúa, J. (2005). "Entre el consenso y la evidencia científica", *Gaceta Sanitaria*, 19(1).

És freqüent que molts dels temes sobre els quals es vol incidir amb un programa manquin d'estudis suficients que permetin fer una revisió sistemàtica mínimament rigorosa. És en situacions d'aquest tipus en què cal buscar alternatives a la mera decisió subjectiva i discrecional. Entre aquestes, la proposta de buscar un consens d'experts qualificats que debatin i elaborin conclusions sobre la base de les evidències empíriques que puguin aportar és una opció vàlida. El text d'Asúa explica de manera detallada com operativitzar una conferència que coordini grups de treball amb aquesta finalitat.

Fernández, S. i altres (2002). "Evaluación de la efectividad de los programas escolares de prevención del consumo de tabaco, alcohol y cannabis: ¿qué nos dicen los meta-análisis?". *Revista Española de Salud Pública* (vol. 76, pàg. 175-187).

1) Volem cridar l'atenció sobre l'existència de bases específiques de publicacions, en aquest cas Cochrane i Medline, i de la necessitat de triar descriptors o paraules clau eficaços que ens permetin seleccionar de manera exhaustiva però precisa les publicacions.

2) Volem destacar també que l'article no és una metanàlisi en si mateixa, sinó una revisió sistemàtica de metanàlisis prèvies. Aquesta opció no s'ha explicat en el text del mòdul, i convé que l'estudiant ho tingui en consideració.

3) Observeu que la magnitud de l'efecte dels programes ha estat estimada en unitats de desviacions estàndard (taula 2).

4) La part del text enfocada a descriure les limitacions dels estudis consultats ha d'ajudar l'estudiant a desenvolupar la capacitat de crítica per a jutjar la validesa de la metanàlisi a partir dels cinc criteris de qualitat de Bruvold.

Ivàlua (2011). *Avaluació ex-ante. Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques*.

Aquesta guia és un molt bon complement al text del mòdul. Reforça i amplia conceptes, a la vegada que relaciona l'avaluació *ex ante* amb la teoria de la causalitat i de l'explicació científica que s'ha vist en el mòdul "Els reptes prin-

cipals d'una avaluació d'impacte". És a més molt necessari que l'estudiant es fixi en els passos que s'han de seguir i per què, abans que es faci cap recomanació sobre l'oportunitat i la configuració d'un programa.

Meseguer, F. (2007) "Lectura crítica de un Metaanálisis y de una revisión sistemática", Consejería de Sanidad, *Atención sanitaria basada en la evidencia: su aplicación a la práctica clínica*, Consejería de Sanidad, Murcia (capítol 11).

Disponible a:

http://www.murciasalud.es/publicaciones.php?op=mostrar_publicacion&id=103&idsec=88

1) L'anàlisi crítica de les revisions sistemàtiques i les metanàlisis està ben presentada en aquesta lectura en què l'autora emfatitza els criteris de qualitat mínims que ha de complir aquesta mena de recerca.

2) L'estudiant ha de trobar en aquesta lectura una guia pràctica sobre com ha d'afrontar una revisió sistemàtica i una metanàlisi.

3) Hi trobarà també un glossari molt complet que és complementari del que ja ha vist en començar l'assignatura, que està enfocat en l'avaluació, mentre que aquest està més focalitzat en les revisions sistemàtiques i metanàlisis.

4) És interessant que presti atenció a les condicions sota les quals no té sentit fer una metanàlisi.

Altres lectures recomanades

Higgins, JPT, Green, S. (editors). *Manual Cochrane de revisiones sistemáticas de intervenciones*. Traducción del Centro Cochrane Iberoamericano de: *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

http://www.cochrane.es/files/handbookcast/Manual_Cochrane_510.pdf

Molinero, L. M. (2003). *Meta-análisis*.

Disponible a: <http://www.seh-lelha.org/metaanalisis.htm>