

Els reptes principals d'una avaluació d'impacte

Sebastià Sarasa Urdiola

PID_00197901



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

1. Què significa avaluar l'impacte d'un programa? Teoria, causalitat i hipòtesi.....	5
2. Definir el contrafàctic.....	11
3. Utilitzar la comparació amb un altre grup com a contrafàctic.....	13
4. Mètode reflexiu de construir un contrafàctic.....	15
5. Biaix d'endogenitat i efecte selecció.....	16
6. Interferència de factors externs.....	19
7. Validesa de l'avaluació.....	20
8. Lectures complementàries.....	21

1. Què significa avaluar l'impacte d'un programa? Teoria, causalitat i hipòtesi

Abans de quantificar l'efecte que té un programa, cal demostrar que el programa que avaluem és la **causa** d'un efecte específic. En ciències socials la definició d'una relació causal entre dos fets es fa sempre en termes de probabilitats que un esdeveniment es produeixi, mai en termes de seguretat absoluta. Es pot agafar un got d'una taula i deixar-lo anar al buit amb la certesa absoluta que el got caurà immediatament a terra, a causa de la força de la gravetat, però en ciències socials, si diem que **A és causa de B**, només estem assegurant que **si es produeix A, la probabilitat que succeeixi B és més elevada**.

Pensem, per un moment, en la relació causal que hi pugui haver entre atur i delinqüència. No totes les persones que han perdut la feina esdevenen delinqüents, tampoc no tots els delinqüents són persones que no tenen cap feina legal, per tant, no podem assegurar amb total certesa que si una persona és acomiadada, tard o d'hora cometrà un delicte. Però és un fet empíricament demostrat que les persones sense ocupació legal remunerada **tenen més probabilitats** de cometre cert tipus de delictes, i també que les xifres de delictes comesos en un país **augmenten amb més probabilitat** en períodes de recessió econòmica i augment de l'atur. És en aquest sentit que podem afirmar que l'atur és **una causa** de la delinqüència. I diem una causa, perquè no és l'única. Per la mateixa raó, un programa d'ocupació juvenil, que tingui com a efecte esperat una reducció de la delinqüència entre els joves, l'hauré d'avaluar mesurant si entre els participants en el programa hi ha hagut **una reducció de la probabilitat** de delinquir en comparació amb la probabilitat que tenen els que no hi han participat.

L'avaluador farà bé a no oblidar que una teoria és una visió parcial de la realitat i que la causalitat en ciències socials és en termes de probabilitats. De vegades, això és difícil d'entendre per a algunes persones que, perquè es deixen endur per fets anecdòtics o bé perquè tenen dificultats per a entendre la causalitat com una relació de probabilitats, asseguren que un programa fracassa perquè coneixen un o més casos d'individus que, malgrat haver participat en el programa, no han mostrat els efectes positius esperats.

L'avaluació de l'impacte d'un programa pretén validar si una intervenció organitzada amb una finalitat té efectes significatius en la consecució d'aquesta finalitat. Traduït al llenguatge de les ciències socials, direm que es pretén contrastar que entre la variable explicativa (la intervenció) i la variable explicada (la finalitat) hi ha una associació estadística, és a dir, que variacions en la variable explicativa es tradueixen en variacions del mateix signe, o de signe con-

Bibliografia

R. S. Johnson; S. Kantor; P. V. Fishback (2007). *Striking at the Roots of Crime: The Impact of Social Welfare Spending on Crime During the Great Depression*, Working Paper 12825 National Bureau of Economic Research.

trari, en la variable explicada. A més, que aquesta associació entre les variables no és espúria, i que hi ha un model teòric raonable i versemblant que dona suport a l'associació observada (Lago, 2008).

L'avaluador de programes ha de tenir clar que un programa ha d'estar sustentat en una teoria que fonamenti una explicació causal, sobre la qual s'ha bastit una hipòtesi sobre l'efecte esperat de la intervenció (Birckmayer i Weiss, 2000). Tots aquests conceptes, *teoria*, *explicació causal* i *hipòtesi*, estan relacionats, però són diferents d'allò que és una mera descripció dels fets. Les descripcions ens mostren uns fets que poden tenir, o no, alguna associació entre ells. Per exemple, si constatem que hi ha una associació estadísticament significativa entre les xifres de persones que estan a l'atur i la percepció subjectiva d'inseguretat ciutadana, estem fent una mera descripció del fet però no estem donant cap explicació del perquè hi ha aquesta associació, és més, podria donar-se el cas que aquesta relació entre les variables atur i percepció d'inseguretat ciutadana fos una **relació espúria** (diem que una relació entre variables és espúria quan no obeeix a cap procés directe d'influència d'una sobre l'altra, sinó que és una relació purament casual, o una relació induïda per una tercera variable que no podem observar). Per tant, un cop observada aquesta associació hem de fer-nos preguntes del tipus: hi ha una relació causa-efecte entre totes dues variables? I si la resposta és sí, com expliquem aquesta relació?

Per a comprovar que hi ha una relació causa-efecte hem d'analitzar les covariacions entre les dues variables i d'altres que puguin tenir algun efecte sobre elles. En termes estrictament estadístics, direm que entre dues variables hi ha una explicació causal quan, aprofitant les eines de l'anàlisi multivariable, i amb un model de regressió ben especificat, podem rebutjar la hipòtesi nul·la que no hi ha relació entre elles, un cop les altres variables que puguin alterar aquesta relació han estat controlades. Dit d'una altra manera, quan mantenint totes les altres variables constants, una variació de la variable explicativa es tradueix en una variació de la variable explicada. Obtenim més seguretat que hi ha una relació causal entre dues variables quan treballem amb dades organitzades en sèries temporals que permetin observar com una variació de la variable explicativa en el moment t provoca una variació de la variable explicada en el moment $t + n$ (Blossfeld i Rohwer, 1995).

En suma, necessitem precisar amb cura quina és la relació causal estadística que hi ha entre les variables que hem observat. La relació entre dues variables X i Y pot prendre diverses formes, com ens mostra la taula 1.

Taula 1

- a) $X \rightarrow Y$
- b) $Y \rightarrow X$
- c) $X \leftarrow \rightarrow Y$
- d) $X \rightarrow YZ = 1$

e) $Z \rightarrow X \rightarrow Y$
 $Z \rightarrow Y$

f) $Z \rightarrow X$
 $Z \rightarrow Y$

g) $X \neq Y$

En l'opció *a* de la taula, la variable *X* és causa de la variable *Y*.

En l'opció *b* la relació és inversa, *Y* és causa de *X*.

L'opció *c* ens mostra una situació anomenada *d'endogenitat*, les dues variables es modifiquen mútuament de tal manera que *X* causa *Y*, a la vegada que *Y* causa *X*.

Podrem veure més exemples d'endogenitat quan abordem el mètode d'avaluació amb variables instrumentals. Hi ha també endogenitat en el cas *e*, en què la influència de *X* sobre *Y* està influïda per una altra variable *Z* que afecta les dues variables, l'explicativa i l'explicada. Aquest és un exemple de mecanisme mediador, del qual parlarem més endavant.

El cas *d* és una representació gràfica de l'anomenat *mecanisme moderador*, del qual també parlarem més avall, i que hem de llegir com que *X* influeix sobre *Y* només si es dóna la condició *Z*.

Per acabar, ens queden les opcions *f* i *g*. En aquestes dues opcions la relació entre les variables *X* i *Y* és espúria. En el primer cas (*f*) perquè la covariació observada entre les dues no es deu al fet que mantinguin cap relació causal, sinó al fet que totes dues estan causades simultàniament per una tercera variable *Z*. En el segon cas (*g*), simplement no hi ha cap relació regular entre *X* i *Y*, i podríem pensar que la relació que nosaltres hem observat és fruit de la casualitat, o d'un error en la mesura.

Cal, però, distingir entre l'explicació estadística i l'explicació causal.

Prenem com a exemple l'associació que hi ha entre la renda mitjana *per capita* d'un país i la privació relativa que tenen algunes llars en el consum de béns bàsics com ara la impossibilitat de menjar amb freqüència diària carn o peix, de poder fer front a despeses imprevistes, o de gaudir almenys d'una setmana de vacances a l'any fora de casa. A grans trets, hi ha una associació estadísticament negativa entre totes dues variables, ja que als països més rics, amb una renda *per capita* elevada, hi ha menys proporció de llars que pateixen aquestes privacions. Podem inferir doncs que un augment en la riquesa de les nacions ha de comportar una reducció de la privació en el consum de béns bàsics i,

Oferta i demanda

Un dels exemples més citats d'endogenitat és la fixació del preu d'un bé o servei com a resultat de la interacció entre oferta i demanda. Si el preu d'un bé baixa, n'augmenta la demanda, però l'augment de la demanda i la retirada d'una part de l'oferta a la qual ja no surt rendible el nou preu, fan pujar de nou el preu. De tal manera que preu, volum d'oferta i volum de demanda es defineixen de manera endògena, sense que poguem dir, utilitzant una dita popular, què és primer, si l'ou o la gallina.

fins i tot, els organismes internacionals faran bé de recomanar que les nacions procurin fer créixer el producte interior brut com a estratègia per a reduir el volum de població afectada per la privació en el consum de béns bàsics.

Però, sense més informació addicional, l'associació que observem entre privació i renda *per capita* **no és una explicació causal**. Una explicació requereix una teoria que faci una descripció analítica de com la variació en la variable explicativa influeix en la variable explicada (o, en el llenguatge en excés mecanicista d'alguns textos, quin és el mecanisme que uneix la dinàmica de les variables explicativa i explicada). En l'exemple que hem triat, una possible explicació seria que, en augmentar la riquesa col·lectiva d'una nació, augmenta la demanda de béns i serveis de manera que també creix l'oferta de llocs de treball, tant qualificats com no qualificats, i, per tant, les llars de tots els estrats socials gaudeixen d'un augment de renda, incloent-hi les més pobres, i seria així com més persones podrien consumir béns bàsics que abans no podien. Quan observem l'associació estadística entre renda mitjana *per capita* i privació no observem directament el procés explicatiu, assumim que l'augment de la renda mitjana beneficia a tothom, i diem que les dades observades són **consistents** amb la teoria que un augment de la riquesa mitjana redueix la privació. Si l'associació entre aquestes dues variables no existís, o fos en sentit invers a l'esperat, diríem que la relació observada entre renda *per capita* i privació no seria consistent amb la nostra teoria.

Però per a contrastar empíricament una explicació necessitem quelcom més que trobar dades indirectes que són consistents amb la teoria. En el cas que veiem, necessitem demostrar que, en efecte, un augment de la renda mitjana es tradueix en un augment de la renda dels més desfavorits, que tenen més risc de patir privacions, i que aquest augment de renda entre els més desfavorits no és anul·lat per un creixement en el cost dels béns de consum que anul·li la major renda disponible. Si això es complís sempre, tindríem una teoria sobre el procés pel qual el creixement de la renda redueix la privació relativa. Ara bé, podria donar-se el cas que en alguns països, o en alguns moments històrics, no hi hagués cap associació entre creixement de la renda mitjana i privació de béns bàsics, la qual cosa es podria explicar per l'existència de coalicions polítiques, o situacions d'oligopoli, que desviessin el creixement de renda en benefici exclusiu d'uns grups socials, exclouent els més desfavorits. Llavors, la teoria que explica la reducció de la privació pel creixement de la riquesa nacional mitjana no seria del tot correcta, i seria més encertat reformular-la dient que els augments de renda mitjana redueixen la privació dels grups més desfavorits si hi ha condicions que afavoreixin el flux de rendes addicionals cap a aquest grups.

Tornant al llenguatge mecanicista, diríem que hi ha un **mecanisme moderador** (vegeu la situació *d* de la taula 1) que condiciona l'efecte de la variable explicativa sobre la variable explicada. Llavors el factor causal seria l'augment

mitjà de la renda, però condicionat per la capacitat dels diferents grups socials per a controlar el destí d'aquests augments. Així doncs, els mecanismes moderadors són les condicions sota les quals el mecanisme explicatiu operen.

L'exemple anterior ens il·lustra la importància que tenen les condicions sota les quals una hipòtesi s'acompleix, i que s'han de tenir en consideració tant en el moment de dissenyar el programa com en el moment d'avaluar-lo.

Exemple

Prenem un altre exemple: el problema de la violència masclista contra les dones. A grans trets tenim tres teories en disputa sobre les causes d'aquesta violència. Una teoria apel·la a condicions estructurals i culturals macro i addueix que vivim en societats patriarcals, on els homes exerceixen el domini sobre les dones amb diversos recursos de poder, i entre ells, la violència física (Dobash i Dobash, 1979). Les altres dues teories, sense negar la raó a la primera, adverteixen que, per bé que totes les societats actuals conegudes se sustenten en una base patriarcal, mostren graus de violència masclista significativament diferents i miren de donar raó d'aquestes diferències. La teoria més racionalista declara que l'home i la dona negocien en l'esfera privada les condicions de convivència i el resultat de la negociació dependrà dels recursos de poder que cadascú tingui (Brines, 1994). Recursos que depenen molt del context social en què estan immersos. Així, la capacitat dels homes per a exercir violència física sobre les dones serà inferior en els països i estrats socials en què les dones tinguin més oportunitats d'obtenir ingressos del treball pel seu compte i puguin amenaçar els marits amb el divorci. Com a prova mostren que la violència és més elevada en societats i estrats socials en què la dona té menys oportunitats laborals i menys oportunitats de viure per si mateixa. L'altra teoria apel·la més al caire irracional de la violència masclista, i addueix que la violència és més freqüent quan l'home pateix frustracions severes que no sap gestionar de manera saludable, i que es tradueixen en conductes rabioses envers l'entorn i, en especial, envers els éssers més pròxims i febles. Una de les circumstàncies en les quals els homes poden viure una intensa frustració es dona quan perden el paper de sustentador principal de la llar, perquè ell perd la feina o simplement perquè la dona té una carrera laboral més prometedora que ell (Gelles, 1974). Com a prova empírica a favor d'aquesta teoria es mostra una dada que hi és consistent: l'elevada taxa de suïcidis entre els homes que maten les seves dones, la qual cosa mostraria la irracionalitat de la seva conducta. Tenim, doncs, tres teories, tres interpretacions raonables i versemblants, que identifiquen causes diferents del mateix fet.

Imaginem ara que en un país fictici de la Unió Europea, les autoritats comproven amb les estadístiques judicials i policials que la violència masclista afecta de manera significativa més les dones d'estrats socials humils, en la majoria immigrants de països més pobres, on les relacions familiars es basen en pautes tradicionals de més submissió femenina, i que estan casades amb homes també immigrants amb elevada precarietat laboral i risc d'exclusió. El dilema per a les autoritats a l'hora de dissenyar un programa preventiu és enorme, ja que les dades disponibles s'ajusten per igual a les tres teories i no saben a quina donar més importància per a dissenyar els recursos necessaris.

Suposem que les autoritats consideren que, des de l'enfocament de la teoria del patriarcat, un programa preventiu de la violència masclista hauria d'anar adreçat a revolucionar totes les institucions socials que fonamenten la nostra societat, amb efectes a llarg termini, i les autoritats prefereixen optar per prioritzar una estratègia que pugui donar resultats positius a més curt termini. Suposem també que els encarregats d'oferir una alternativa consideren raonables les dues teories restants i opten per fer un programa que consideri la interacció entre els dos processos causals que se'n deriven. Augmentar els recursos de poder de les dones promovent-ne l'ocupació laboral, serveis de cura dels infants i habitatge assequible en cas de voler la separació del marit, i a la vegada un programa de suport personal i psicològic, i de mediació, en els casos de conflictes interpersonals que ajudi a contenir la frustració dels homes. Tot acceptant les hipòtesis de partida següents:

- a) Si augmenta l'ocupació femenina, condicionada a uns salaris i uns costos de l'habitatge i de la cura dels menors que permetin a les dones una vida autònoma en cas de necessitat, augmenta la capacitat de les dones de dissuadir les parelles de tenir comportaments violents.
- b) Si augmenta l'ocupació femenina en els estrats socials més humils, augmenta la renda disponible de les llars i es redueix l'estrès derivat de la pobresa que alimenta la frustració dels homes.

c) Un programa de suport personal i de mediació en els conflictes de parella que faci reflexionar els homes sobre les bases de la seva identitat i els avantatges de l'ocupació de les seves dones s'ha de traduir en un descens de la frustració entre homes sense ocupació remunerada, tot condicionat al fet que s'hagin complert també les hipòtesis *a* i *b*.

d) Les circumstàncies esmentades han de reduir el risc que tenen les dones de llars més humils de patir violència masclista.

En suma, insistim que una explicació estadística no equival a una explicació teòrica; i que l'explicació teòrica ha de ser capaç d'oferir una resposta raonada a la pregunta de si hi ha cap mena de relació entre dos fenòmens observats. Tant si es tracta d'explicar el curs d'una malaltia, la modificació en els trets de personalitat d'un individu, o la conducta esperada dels habitants d'un barri o d'un país, una teoria és una especulació que ofereix un argument raonable i versemblant sobre el curs observat i esperat dels esdeveniments. Una explicació teòrica requereix, doncs, que hi hagi un seguit de proposicions predictives que poden ser traduïdes a hipòtesis sobre quin seria l'efecte d'un tractament en un experiment, i quines són les condicions contrastables que s'han de complir perquè l'efecte sigui l'esperat.

Cal tenir present que un programa tindrà deficiències en el disseny, i dificultats en l'avaluació, si els responsables del disseny i avaluació no són capaços d'explicar els **mecanismes moderadors**, és a dir, sota quines condicions el programa tindrà els efectes positius esperats, i sota quines condicions el programa no tindrà efectes o tindrà efectes contraproductius. La resposta a aquestes preguntes requereix tenir una teoria sòlida sobre el problema que volem abordar amb el programa.

Aquesta teoria també ens ha de permetre identificar els possibles mecanismes mediadors. Un **mecanisme mediador** és aquell, mitjançant el qual, la variable explicativa influeix la variable explicada quan aquesta influència no és directa, sinó mitjançant una tercera variable (vegeu la situació *e* de la taula 1). Aquesta mena de relació és força freqüent en les ciències socials.

Posició social

Un exemple de mecanisme mediador el trobem en com es transmet la posició social de pares a fills en societats no agrícoles. A la pràctica totalitat de les societats modernes, hi ha una correlació elevada entre les posicions socials de pares i fills, però el mecanisme que transmet la posició és, entre d'altres, les credencials educatives que obtenen els fills, les quals condicionen la posició en el mercat de treball. Atès que el rendiment educatiu dels joves està condicionat per la classe social dels pares, és mitjançant la desigualtat educativa que es transmet la desigualtat social entre generacions. En aquest cas, l'educació és el mecanisme mediador que explica part de la influència que exerceix la posició social dels pares sobre la posició social dels fills.

2. Definir el contrafàctic

Volem saber el resultat que ha tingut un programa en la població objectiu a la qual va adreçat. Assumim que la població realment usuària del programa és la que hauria de ser perquè hi ha hagut una **avaluació de seguiment** prèvia que ens confirma que les desviacions entre els perfils d'usuaris previstos i els reals han estat insignificants. Ara el problema rau a saber si el programa ha tingut cap efecte en les vides de les persones i si aquest ha estat positiu o negatiu. La ciència experimental valida els efectes causals comparant què succeeix amb un material quan està sotmès a unes condicions i quan no està subjecte a les mateixes condicions. En les ciències socials, però, no és possible observar un mateix individu sota condicions de vida alternatives estrictament independents una de l'altra.

Imaginem, per exemple, un individu anomenat Joan que obté una feina en una central nuclear en un moment determinat del seu curs vital, quan té vint-i-dos anys, i acaba desenvolupant, deu anys més tard, actituds favorables a aquesta mena d'energia que abans no tenia. Podem intuir que el canvi ha estat fruit de la nova ocupació, però mai no podrem observar com pensaria aquest mateix individu als trenta-dos anys d'edat si mai no hagués treballat en una central nuclear.

Aquesta limitació obliga els investigadors socials a tenir molta cura en l'ús dels anomenats *contrafàctics*, és a dir, les situacions que no han succeït però sobre les quals establim hipòtesis dels seus efectes en cas que haguessin succeït (Gertler i altres, 2011). En el cas esmentat, els **fets** són que Joan va començar a treballar en una central nuclear als vint-i-dos anys quan no tenia una actitud favorable a l'energia nuclear i deu anys més tard ha canviat d'actitud. El contrafàctic (allò que podria haver passat però mai no va succeir) és que Joan no hauria obtingut aquesta feina en la central nuclear. Llavors, l'avaluació de l'impacte que ha tingut l'ocupació laboral en els judicis de l'individu es fa comparant els fets reals amb uns altres fets possibles, però que no han succeït. És a dir, el repte, doncs, és com definim el contrafàctic.

En el cas de l'avaluació de l'impacte d'un programa el repte és identificar que hauria passat si el programa no hagués existit, i el contrafàctic és la situació que tindrien els usuaris del programa si la intervenció no hagués tingut lloc. Però, com que els individus no podem tenir dues existències simultànies, els avaluadors estan obligats a construir grups d'individus amb qui comparar els que han estat subjectes a la intervenció del programa, i la selecció del grup de comparació ha de ser prou raonable i convincent perquè aquest grup sigui acceptat com a contrafàctic, és a dir, equivalent al grup d'individus que ha participat en el programa.

La manera de construir el contrafàctic varia segons les circumstàncies i la disponibilitat d'informació, però a grans trets podem distingir dos mètodes: la comparació amb un **grup de control** diferent de l'experimental, i el **mètode reflexiu**, que compara els resultats del grup experimental abans i després de la intervenció, sense comparar-lo amb cap altre grup més.

3. Utilitzar la comparació amb un altre grup com a contrafàctic

Comparar els resultats dels participants en el programa amb els obtinguts per persones que no hi han participat és una alternativa, però aquest mètode és incorrecte quan hi ha un **efecte selecció** en els integrants del programa que els fa diferents dels individus del grup de comparació.

Exemple

Suposem que hi ha un programa d'atenció a infants que cursen la primària i que viuen en condicions de pobresa. Suposem que l'objectiu del programa és millorar-ne el rendiment escolar i, en el moment d'iniciar els estudis d'ESO, comparant els resultats obtinguts pels participants en el programa amb els obtinguts per la resta d'estudiants que no hi han participat, es descobreix que el rendiment dels participants en el programa està per sota de la resta.

Vol dir això que el programa és inútil, fins i tot perjudicial, i que s'ha de tancar? La resposta és no. No, fins que no es faci una avaluació més rigorosa que compari nens vivint sota les mateixes condicions. Els nens que viuen en la pobresa tenen de mitjana un rendiment escolar per sota de la mitjana de la població escolar i és possible que, malgrat que el programa en millori el rendiment, aquest encara estigui per sota de la mitjana. Si aquest fos el cas, el programa no hauria servit per a superar totes les dificultats que pateixen aquests nens, però encara estaria per demostrar que, sense el suport del programa, el rendiment escolar no hauria estat pitjor.

Hem de tenir al cap que l'avaluació dels efectes d'un programa es fa comparant els resultats del grup que hi ha participat amb els resultats que el mateix grup hauria obtingut en cas que no ho hagués fet (el contrafàctic), i que hem de buscar un grup de comparació de característiques molt semblants a les que tenen els participants en el programa. En aquest cas, l'avaluació correcta compararia els nens de la mateixa edat que participen en el programa amb els que no hi participen, però vivint tots en condicions similars.

El repte de definir el contrafàctic és així el repte d'identificar quins individus, no participants en el programa, tenen característiques similars a les dels participants. En aquesta tasca ens ajuda la teoria i la recerca prèvies que s'hagin fet en el camp que ens interessa, i que hagin elaborat models causals contrastats empíricament sobre els factors que estan associats a la nostra variable dependent. Si, abans de posar en marxa el programa, s'ha fet una **revisió sistemàtica** de les teories que expliquen el problema que volem abordar amb el programa, els resultats seran ara d'una utilitat impagable, perquè ens hauran informat dels factors que cal tenir en compte a l'hora de predir com poden afectar les característiques dels individus, i els seus contextos, en el resultat del programa. En l'exemple anterior, una recerca prèvia sobre rendiment escolar ens hauria informat que aquest està condicionat per factors personals (el sexe, l'estat de salut i la llengua materna dels nens), familiars (la renda disponible, el capital cultural, la formació i el valor que donen a l'educació els pares) i de l'entorn dels estudiants (la composició social del barri i de l'escola concreta on estudia l'individu). Doncs bé, definir el contrafàctic, i establir el grup de com-

paració corresponent, significa construir un grup d'estudiants, no participants en el programa, que tinguin les mateixes característiques personals, familiars i d'entorn que els participants.

4. Mètode reflexiu de construir un contrafàctic

El mètode reflexiu comptabilitza el nombre de participants en el programa que ha experimentat canvis favorables en la seva situació i atribueixen aquests canvis a les bondats del programa.

D'exemples d'aquesta mena n'hi ha molts, com ara programes de formació i ocupació a les presons que mesuren el seu èxit pel nombre d'interns que aconsegueixen una feina en acabar la condemna; programes d'ajuda a la infància que mesuren el seu èxit pel nombre de nens que no abandonen els estudis, etc.

Quan un estudi vol avaluar l'impacte d'una política pública i es fa aprofitant dades d'enquestes al total de la població, el mètode reflexiu exigeix observar els mateixos individus enquestats al llarg del temps, de manera que podem comparar-ne les conductes i actituds abans i després de participar en un programa. L'accés a aquesta mena d'informació és possible quan tenim enquestes tipus panel que fan un seguiment dels mateixos individus al llarg d'un seguit d'anys, com ara l'Enquesta de condicions de vida de la Unió Europea, o el Panel de desigualtats socials a Catalunya que elabora la Fundació Jaume Bofill. També podem accedir indirectament a aquesta informació amb enquestes en què es fan preguntes retrospectives sobre comportaments en el passat. L'inconvenient d'aquestes bases de dades esmentades és que són poques i, a més, amb mostres relativament petites que si bé poden ser útils quan es tracta d'avaluar els efectes de polítiques públiques en el conjunt de la població no serveixen per a fer l'avaluació de programes que afecten grups petits d'individus, ja que cap o pocs d'ells estaran recollits en la mostra.

El mètode reflexiu és menys fiable que el mètode experimental i que el quasiexperimental, en què hi ha un grup de comparació, ja que té més risc de tenir poca validesa interna, en ser incapaç de comparar l'evolució dels resultats entre els participants amb cap altre grup que pugui operar com a contrafàctic. Aquesta mancança impedeix que pugui donar una resposta fiable a la pregunta cabdal: els resultats que obtenen els individus, són realment deguts al programa o haurien obtingut els mateixos resultats sense la intervenció?

Vegeu també

Per aprofundir en el concepte de validesa interna, vegeu l'apartat "Validesa de l'avaluació" d'aquest mòdul didàctic.

5. Biaix d'endogenitat i efecte selecció

Un **biaix** en l'estimació del valor d'una variable és una desviació del seu valor real que es dona de manera consistent, és a dir, que si repetim el mesurament amb altres individus, la desviació es torna a repetir. El biaix d'una estimació és diferent de l'**error aleatori** que es dona en tot mesurament, i que no segueix una pauta previsible. Entendrem millor la diferència entre biaix i error aleatori si pensem en dos mètodes per a seleccionar els participants en un programa.

Exemple

Imaginem que tots dos fan una selecció a l'atzar dels participants. Un, a partir del padró municipal, selecciona per sorteig una petita mostra de trenta individus i comprova que entre els seleccionats hi ha vint homes i deu dones. Aquesta proporció no es correspon amb la distribució de la població per sexe i és atribuïble a l'error mostral, amb el benentès que aquest error tendeix a desaparèixer a mesura que augmentem la grandària de la mostra. Segurament, una nova selecció de trenta individus més donarà una altra composició per sexe, i a mesura que sumem extraccions de mostra la proporció d'homes i dones tendirà a l'equilibri poblacional.

Imaginem ara que l'altre mètode de selecció dels participants consisteix en un mostreig aleatori entre els socis dels clubs de futbol, i observem que per més gran que fem la mostra sempre hi ha una majoria d'homes entre els seleccionats. En aquest cas no es tracta de cap error aleatori, sinó del fet que la població de la qual extraïem els participants és una població amb un biaix de composició a favor dels homes, ja que els socis dels clubs de futbol són majoritàriament homes.

Diem que hi ha **endogenitat** quan es dona alguna de les circumstàncies següents:

1) La variable dependent (el resultat del programa) i una variable independent (explicativa del resultat del programa) mantenen una relació bidireccional d'influència mútua. Aquest seria el cas *c* que il·lustra la taula 1.

Per exemple, en el cas d'un programa que pretengui promoure activitats de lleure i esport tutelades per a prevenir el consum de drogues, hem de tenir en compte l'endogenitat de totes dues variables que mantenen una relació bidireccional. D'una banda, els individus que tenen moltes hores desocupades i no gaudeixen de cap activitat organitzada de lleure tenen un risc més elevat de consumir drogues, però, al mateix temps, els individus que han iniciat un procés de drogodependència tendeixen a abandonar les pràctiques esportives i de lleure organitzat.

2) Hi ha una variable explicativa no observada que es correlaciona amb alguna de les observades, seria el cas *f* de la taula 1.

Per exemple, hi ha risc d'endogenitat quan la participació en el programa és voluntària i l'avaluació de l'impacte la fem comparant els resultats dels participants amb un grup format per persones que no han estat interessades a participar-hi. En certa manera, es tractaria d'un **efecte d'autoselecció** (vegeu més avall), ja que formar part del programa podria obeir a uns factors associats directament al resultat obtingut. Per exemple, els participants en un curs de formació específic poden tenir tanta voluntat per a millorar la seva posició laboral que, amb independència del curs, els ajudarà a pujar d'estatus. Com que la voluntat dels individus és difícil de mesurar, tindrem una variable explicativa no observada que està directament relacionada amb el resultat del programa i amb la probabilitat que els individus hi participin. Això es tradueix en el fet que les nostres estimacions atribuiran a la participació en el programa de formació uns efectes positius sobre la carrera laboral que recolliran els efectes de la voluntat i, per tant, no podrem estar segurs de si aquests individus no haurien obtingut els mateixos resultats encara que mai no haguessin rebut la formació del programa.

Hem esmentat que una de les causes d'endogenitat pot derivar de l'**efecte selecció**, però en què consisteix aquest efecte? A grans trets diem que l'estimació dels efectes d'un programa està esbiaixada a causa de l'efecte selecció quan hi ha hagut un error en la selecció dels individus que formen part del programa o del grup de control, de tal manera que aquests no són homogenis entre ells i, per tant, no són comparables.

L'efecte selecció pot ser degut a causes diferents:

1) Defectes en la selecció de la mostra que deriven en una composició d'individus que no representen el conjunt de la població.

Per exemple, els enquestadors encarregats de fer les entrevistes que han de recollir informació sobre conductes dels adolescents poden tenir por d'entrar en certs blocs d'habitatges en barris marginals, de tal manera que tendeixin a evitar-los, esbiaixant així la representativitat de la mostra. Aquesta mena de biaix és difícil de corregir si els responsables de l'avaluació no són conscients que s'ha produït. En cas de descobrir-ho, el remei pot ser forçar més entrevistes dels segments infrarepresentats o, si ja és tard o el pressupost no ho permet, ponderar la mostra de manera que els casos infrarepresentats tinguin més pes.

2) Biaix de susceptibilitat, quan el grup que participa en el programa té alguna característica que el fa especialment susceptible davant el tractament i, en conseqüència, la seva resposta al programa no serà representativa de la resposta mitjana de la població.

Per exemple, els infants de llars pobres tenen més risc de fracàs escolar, la qual cosa fa que, si focalitzem un programa de reforç escolar en nens pobres i després en comparem el rendiment escolar amb els resultats que obté el conjunt de la població infantil, probablement els resultats seran inferiors. Encara que això no significa que sense el programa no haurien estat pitjors.

3) El biaix d'autoselecció dels participants en el programa és una variant del biaix de susceptibilitat, però si en aquest cas es dona no és perquè els responsables del programa hagin seleccionat una població de risc específica, sinó perquè la participació en el programa és voluntària i hi ha factors, observats o no observats, que estan correlacionats amb el fet que un individu prengui la decisió de participar-hi i, a la vegada, amb el resultat del programa. Si aquests factors no són corregits, l'estimació que farem de l'impacte del programa serà errònia.

4) El biaix pel desgast de la mostra. Quan l'avaluació d'impacte es fa amb un seguiment d'individus al llarg del temps, cal tenir en consideració que, a mesura que transcorre el temps, molts individus abandonen el grup de tractament i el grup de control per raons diverses: alguns canvien de domicili sense comunicar-ho als avaluadors, i d'altres abandonen per cansament. Hi ha la possibilitat de reduir el desgast si s'estableix un programa d'incentius amb retribucions materials o simbòliques als participants en compensació per la seva contribució, però tot i així és difícil que no hi hagi un cert abandó. Si els individus que abandonen tenen un perfil similar al del conjunt de la població, el desgast no ha de ser preocupant, ja que els individus restants continuen essent representatius del total. Ara bé, és freqüent que l'abandó es distribueixi de manera desigual i que es concentri més en un tipus d'individus que en d'altres. En aquest cas, els resultats de l'avaluació estaran esbiaixats, ja que seran el resultat corresponents als individus supervivents en la mostra, i no seran representatius del conjunt de la població.

6. Interferència de factors externs

Penseu en el cas hipotètic següent.

Un govern local de la perifèria d'una gran ciutat és conscient que bona part dels joves tenen dificultats per a trobar feina i endega un programa de formació professional amb incentius a les empreses locals per a contractar els participants en el programa. A tots els joves que hi participen se'ls passa una enquesta amb preguntes retrospectives per a estimar el nombre d'hores treballades al llarg de l'any anterior al començament del programa. Tres anys més tard es torna a passar l'enquesta als mateixos individus i es comprova que la mitjana d'hores treballades ha augmentat de manera significativa. Els responsables del programa es mostren satisfets, però una veu crítica el mètode reflexiu utilitzat i els diu que està demostrat que la taxa d'ocupació augmenta amb l'edat i que ara els joves que van participar en el programa són tres anys més grans, raó per la qual és probable que els joves igualment haurien augmentat les hores treballades sense necessitat del programa, per tant, la utilitat dels diners públics invertits estaria per demostrar.

En aquest cas, utilitzar el mètode reflexiu d'avaluació ha estat un error, ja que un factor extern, el temps transcorregut, ha alterat les probabilitats que es produeixi el resultat buscat, amb independència de les bondats del programa.

Una altra cosa hauria estat si s'hagués construït un grup de comparació de joves de la mateixa edat que els joves participants en el programa, i si l'edat mitjana dels dos grups s'hagués mantingut similar en el moment de fer l'avaluació. Però si el grup de comparació hagués estat format per joves d'un altre municipi, on al llarg del període observat hi hagués hagut un creixement de l'ocupació generalitzat superior al del municipi del programa, llavors un altre factor extern (la demanda global de força de treball) hauria alterat els resultats.

7. Validesa de l'avaluació

La validesa d'una avaluació fa referència a si l'avaluació mesura realment allò que ha de mesurar (validesa interna); és a dir, si els resultats observats són causats inequívocament pel programa, i si aquests resultats són generalitzables al conjunt de la població i a d'altres contextos (validesa externa).

Per **validesa interna** s'entén que el model estadístic especificat per a explicar l'efecte de la variable independent (el tractament del programa) sobre la variable dependent (el resultat esperat) no deixa sense control qualsevol altra variable que pugui alterar de manera significativa la relació entre totes dues variables dependent i independent. Dit en altres paraules, l'avaluació mancarà de validesa interna si no hem descartat de manera convincent que els resultats aconseguits pels participants en el programa no obeeixen a altres factors aliens al programa, o si els resultats estan esbiaixats per alguna de les raons que hem explicat més amunt.

La **validesa externa** fa referència a la generalització que es pot extreure del programa experimental per a tota la població, cosa que no és sempre possible. Diversos condicionants fan que un experiment amb resultats que gaudeixen d'alta validesa interna no sigui transferible a la resta de la població. En el millor dels casos perquè l'experiment l'estem fent amb una població objectiu específica (per exemple, llars pobres d'una ciutat) i la validesa externa està limitada només a aquesta mena de llars. Si aquest és el cas, l'avaluador ha de ser molt conscient de quines són les fronteres precises de la validesa externa del programa que avalua. Però de vegades la validesa externa és més difícil d'establir si no es fan repeticions de l'experiment en diferents poblacions i circumstàncies.

Un exemple comú el tenim en experiències pilot que s'han dut a terme en centres sanitaris o educatius seleccionats que donen resultats positius en bona mesura per la motivació i entusiasme dels professionals implicats en una experiència nova que saben que està essent objecte d'avaluació.

No és infreqüent que un cop aquestes experiències són generalitzades a tot el sector, l'impacte desaparegui en haver-hi menys implicació de tots els professionals.

En els resultats d'una avaluació hem de garantir preferentment la validesa interna del model explicatiu emprat, tot i que la validesa externa no està garantida per molta cura que tinguem en l'especificació del model.

8. Lectures complementàries

Lectures de l'apartat "Què significa avaluar l'impacte d'un programa? Teoria, causalitat i hipòtesi"

Ato, M.; Vallejo, G. (2011, maig). "Los efectos de terceras variables en la investigación psicológica". *Anales de Psicología* (vol. 27, núm. 2, pàg. 550-561).

En aquest article, l'estudiant pot trobar diferents alternatives de mecanismes explicatius, en especial dels mecanismes mediadors i moderadors descrits en aquest material. Se circumscriu a la psicologia, però el contingut és general per a tota explicació causal a les ciències socials.

Casas Anguita, J. i altres (2003). "La encuesta como técnica de investigación. Elaboración de cuestionarios y tratamiento estadístico de los datos". *Atención Primaria* (vol. 31, núm. 8, pàg. 527-538).

Lago, I. (2008). *La lógica de la explicación en las ciencias sociales. Una introducción metodológica*. Madrid: Alianza Editorial.

El llibre de Lago és una introducció molt planera a la lògica de l'explicació causal de les ciències socials en què, entre altres coses, es fa èmfasi en la diferència que hi ha entre una associació entre variables i una explicació que ha de fer esment a alguna mena de mecanisme.

OECD (2010). *Glosario de los principales términos sobre evaluación y gestión basada en resultados*.

Aquest document és un material de consulta en què l'estudiant pot trobar les definicions dels termes més freqüents en l'avaluació de programes. Té l'avantatge que el glossari, a més d'estar fet en llengua castellana, està en anglès i en francès, la qual cosa ha de facilitar la consulta de termes desconeguts en publicacions fetes en aquestes llengües.

Altres lectures

Bernardi, F. (2006). "Los fundamentos del análisis de la historia de acontecimientos". *Análisis de la historia de acontecimientos* (cap. 1). Madrid: CIS ("Cuadernos Metodológicos", 38).

Aquest capítol és una introducció molt accessible a la tècnica d'anàlisi d'història d'esdeveniments, de gran utilitat quan es disposa d'enquestes tipus panel i l'avaluació d'un programa ha de recórrer al mètode reflexiu; si

bé aquesta tècnica és igualment utilitzable amb mètodes quasiexperimentals. L'estudiant ha d'entendre l'avantatge que té aquesta tècnica enfront de les anàlisis d'enquestes transversals i assimilar conceptes fonamentals com ara *esdeveniment*, *episodi*, *temps continu*, *temps discret* i *durades censurades*. Es pot prescindir de l'epígraf 1.2.2 i següents, tret que els estudiants tinguin un nivell bàsic d'estadística que els permeti entendre'l.

Cobo, E. i altres (2005). "Aspectos metodológicos comunes y específicos de las listas de comprobación". *Medicina Clínica* (vol. 125, núm. 1, pàg. 14-20).

Aquest article focalitzat en la investigació mèdica serveix per a cridar l'atenció sobre el suport que les teories explicatives han de tenir en contrastacions empíriques. També perquè l'estudiant vegi clar que en tot programa que es vol avaluar, com en els fàrmacs i les campanyes de prevenció sanitària, hi ha unes premisses i unes hipòtesis que poden ser més o menys explícites, és a dir, uns mecanismes, i que han de ser avaluats seguint uns criteris de rigor metodològic. En el cas de la recerca clínica aquests criteris han estat fixats en les anomenades *l·listes de control*, però encara no hi ha res de semblant en les ciències socials.

Aquesta lectura ha de servir perquè l'estudiant vegi la diferència entre:

- 1) estudis prospectius i retrospectius,
- 2) transversals i longitudinals,
- 3) error aleatori i error sistemàtic,
- 4) validesa interna i validesa externa.

Echéburua, E.; Fernández-Montalbo, J. (2009). "Evaluación de un programa de tratamiento en prisión de hombres condenados por violencia grave contra la pareja". *International Journal of Clinical and Health Psychology*.

Aquest és un bon exemple d'avaluació reflexiva amb limitacions de validesa interna.

- 1) No hi ha comparació amb cap grup de control, per tant és un cas d'avaluació reflexiva, amb les limitacions que això comporta.
- 2) L'estat de la qüestió que s'ofereix sobre allò que se sap dels efectes de programes similars, cap no s'ha dut a terme dins una presó.
- 3) No s'especifica amb claredat quin ha estat el criteri pel qual s'han seleccionat els grups de participants.
- 4) L'estudiant ha de reflexionar sobre el biaix que es pugui derivar en els resultats pel fet que els participants en el programa siguin voluntaris.

5) És molt important que presteu atenció al fet que l'avaluació no pretén mesurar l'impacte en la reincidència dels delinqüents, sinó que focalitza l'avaluació a observar modificacions del tractament en els mecanismes que expliquen aquest tipus de delictes. Aquest cas ha de servir perquè vegeu que tant important és avaluar els resultats en termes de l'*output* final que espera obtenir un programa (en aquest cas seria reduir la reincidència), com els resultats en termes d'alterar el mecanisme que provoca l'*output* (en aquest cas les percepcions sobre les dones i l'ús de la violència).

Gertler i altres (2011). *La evaluación de impacto en la práctica*. Washington: Banco Internacional de Reconstrucción y Fomento / Banco Mundial.

El capítol 3 ofereix una explicació exemplar del contrafàctic i dels falsos contrafàctics més corrents.

