

# Avaluació d'impacte

F. Xavier Ballart Hernández

PID\_00196155



*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

<b>Introducció</b> .....	5
<b>1. La idea de l'avaluació d'impacte</b> .....	7
1.1. Impacte net i amenaces a la validesa .....	7
1.2. Construir un contrafactual .....	8
1.3. Quan és possible fer una avaluació d'impacte? .....	8
1.4. Impacte, per a qui? .....	10
<b>2. Com es fa?</b> .....	11
2.1. Tres grans aproximacions a l'avaluació d'impacte .....	11
2.2. Experiments reals amb selecció aleatòria prèvia .....	12
2.3. Quasiexperiments .....	13
2.4. Anàlisi multivariant .....	18
<b>3. Lectures obligatòries</b> .....	19
<b>Bibliografia</b> .....	21



## Introducció

L'avaluació d'impacte s'identifica amb l'avaluació *ex post*, que intenta contestar a la pregunta de si la política o el programa és eficaç i contribueix o ha contribuït a atenuar el problema o a donar-hi una solució. Algunes de les investigacions que ha fet el redactor d'aquest mòdul tenen a veure amb aquesta pregunta:

- La prohibició de circular sense casc en motocicleta redueix el nombre de morts i ferits greus?
- La intervenció de brigades de voluntaris o l'ús de determinats mitjans personals i materials redueix el nombre d'hectàrees de bosc cremades?
- El programa de la renda mínima d'inserció contribueix que els seus beneficiaris trobin feina? En quins casos i condicions és més probable que això passi?

Aquest pregunta clau entorn dels efectes de les polítiques es pot plantejar en molts altres casos i en relació amb polítiques i programes públics de tota mena:

- Prohibir fumar en llocs públics redueix l'addicció a fumar? Redueix determinades malalties respiratòries, salva vides? Estalvia diners públics?
- Els grups de recerca que reben ajuts públics experimenten un salt qualitatiu en la seva capacitat per a produir resultats en forma de publicacions, patents o participació en altres projectes finançats per administracions de nivell superior?
- Escolaritzar abans nens d'origen social humil té un efecte en el seu rendiment escolar futur?

L'avaluació d'impacte es proposa respondre a una pregunta que no és fàcil: **Què hauria passat si la intervenció pública no s'hagués produït?** Si no estigués prohibit circular sense casc o sense cinturó de seguretat o sense passar la ITV, si els bombers no apaguessin el foc en el bosc o si no hi hagués brigades de voluntaris, si els que no tenen ingressos no rebessin l'ajut públic ni se'ls imposés cap obligació respecte a la seva activitat per formar-se o trobar feina?

En aquest mòdul, com en els dos anteriors, es fa una presentació deliberadament pràctica (per això no s'hi inclouen referències) en la línia de les guies de l'Ivàlua. El text proposat juntament amb la guia Ivàlua corresponent i l'avaluació que serveix d'il·lustració haurien de permetre a l'estudiant una immersió força completa en el tema objecte del mòdul.

# 1. La idea de l'avaluació d'impacte

## 1.1. Impacte net i amenaces a la validesa

L'avaluació d'impacte és, doncs, la modalitat de l'avaluació que intenta aïllar l'impacte net de la intervenció pública de tots aquells altres factors que afecten el problema, per la qual cosa s'observa una evolució, tant si és en sentit positiu com negatiu. Aquesta pregunta, que és la de la causalitat referida a la intervenció pública, presenta moltes dificultats metodològiques perquè en molts casos els canvis en l'entorn o en les mateixes persones (pel fet de madurar o envellir) expliquen una part més important dels canvis en els problemes que no les polítiques o programes públics.

Per exemple, en el cas del problema de l'atur. En moments d'expansió econòmica, l'economia espanyola va ser capaç de generar oportunitats de treball fins i tot per a persones amb baixos nivells de qualificació. Durant aquests anys, els joves o aturats de llarga durada que participaven en programes de formació en oficis trobaven feina. Però també en trobaven, en la mateixa mesura, les persones de característiques semblants que no van seguir aquests programes. Tan sols s'observaven diferències significatives en el cas de les dones de mitjana edat en endavant. En sentit contrari, en moments de crisi econòmica greu, ni els uns ni els altres aconsegueixen trobar feina i algunes recerques sobre joves apunten que els que participen en programes de formació fins i tot en troben menys que joves de característiques semblants que no passen per aquests programes.

La literatura acadèmica identifica les **amenaces a la validesa interna** amb les raons per les quals l'analista es podria equivocar en afirmar una inferència causal entre una intervenció pública i uns determinats impactes. La principal preocupació en avaluació d'impacte és que els impactes observats no siguin deguts al programa o política sinó a un altre factor o explicació alternativa.

Hi ha moltes amenaces a la validesa interna però les més importants són **la història, la maduració i la selecció**. La primera fa referència a canvis en l'entorn (canvis en l'economia, en el clima, en la societat), en definitiva, a coses que passen en el món al mateix temps que la política. La segona fa referència a canvis en els individus deguts al pas del temps, tant si és perquè creixen físicament o mentalment com perquè envelleixen i perden facultats o salut. La tercera fa referència al fet que els procediments de selecció aleatòria en el cas dels experiments o de creació d'un grup de control en el cas dels quasiexperiments no es faci prou bé i els resultats diferents per a uns i altres s'expliquin pel mètode de selecció o creació dels grups i no per la intervenció pública.

També existeix el concepte de **validesa externa**, que ha rebut molta menys atenció en la literatura malgrat ser tan important o més en alguns casos. La validesa externa fa referència a si es troben uns resultats vàlids per a un pro-

grama en un determinat context, si les conclusions sobre aquest programa es poden generalitzar a altres programes similars que es portin a terme en un altre context, lloc o moment.

Per exemple, en el cas de les polítiques europees és molt important la validesa externa o possibilitat de generalitzar els resultats d'una política ja que sovint es té constància de l'eficàcia d'una determinada intervenció en relació amb un problema, però la qüestió és si fent el mateix tipus d'activitat es resoldrà el mateix problema en un altre país o regió.

## 1.2. Construir un contrafactual

L'avaluació d'impacte es basa en la idea de construir el que s'anomena un **contrafactual**, un escenari que permeti fer una comparació amb el que hauria passat en absència de la intervenció pública. D'entrada, en molts casos és gairebé impossible construir un contrafactual simplement perquè la política o el programa és universal, s'ha portat a terme i no es pot saber què hauria passat en la seva absència.

Per exemple, no sabem què hauria passat si Espanya no hagués implementat el carnet per punts. És una política que s'aplica a tots els conductors, que té efectes en les seves actituds i comportament i en conseqüència en la sinistralitat en les carreteres, que és el que es pretén reduir.

La principal dificultat amb què s'enfronta l'analista consisteix a trobar persones, empreses, barris, regions, etc. de característiques similars a les que han participat en el programa que no hi hagin participat. Això no és fàcil i requereix una certa creativitat.

Per exemple, no és fàcil crear un contrafactual per al carnet per punts llevat que es facin comparacions entre països, però resulta evident que els països (Espanya comparat amb Portugal) tenen característiques diferents pel que fa al comportament dels conductors i la sinistralitat de les carreteres.

No és fàcil tampoc comparar barris que reben subvencions perquè millorin urbanísticament i socialment.

En principi és més factible crear un contrafactual per a persones que participen en un programa ja que és possible controlar aquelles característiques personals que tenen a veure amb la política. Per exemple, en el cas dels aturats, és possible controlar l'edat, el sexe, els estudis previs i l'experiència laboral prèvia. Amb tot, hi ha importants factors psicològics com la motivació que són més difícils de controlar.

## 1.3. Quan és possible fer una avaluació d'impacte?

Com es desprèn de la discussió precedent sobre la creació d'un escenari alternatiu, l'avaluació d'impacte no sempre és possible. Hi ha, però, altres limitacions que dificulten la realització d'una avaluació d'impacte i que la converteixen en un tipus d'avaluació que malgrat ser molt demanada no és fàcil de portar a terme.



En relació amb la intervenció i els objectius:

- La **política o programa** han de ser **clars** en el sentit de quina és la intervenció, les activitats, els productes i els canvis o impactes que es pretenen.
- La intervenció ha de ser **estable**, no ha de ser objecte de modificacions contínues ja que aleshores ja no tenim un programa clar sinó diferents intents o versions poc consolidats.
- Els **impactes s'han de poder deduir de forma bastant directa** dels objectius estratègics. Per exemple, són clars en el cas de la sinistralitat en carretera. No ho són en el cas de la cohesió o integració social.
- En el cas que el programa tingui objectius estratègics complexos, cal valorar si es poden mesurar com, per exemple, en el cas de la cohesió social o de la integració social. Si és possible utilitzar diverses mesures, **cal valorar quines són les mesures més adequades**.

Per exemple, en un programa en què l'objectiu és trobar feina, és suficient un treball informal o ha de ser amb seguretat social? És suficient trobar una feina o mantenir-se en la feina sense perdre-la durant un cert temps? S'hauria de tenir en compte un nivell de retribució mínim?

- Si la intervenció té **més d'un objectiu estratègic**, és imprescindible entrar en la discussió de **quin es mesura**. Per exemple, en les polítiques de rendes mínimes es podria interpretar que hi ha dos objectius: mantenir un nivell mínim d'ingressos que serveixi de coixí per a evitar que la unitat familiar caigui en una situació de pobresa i inserir en el mercat laboral els seus membres perquè siguin financerament autònoms. Avaluat tots dos objectius significa fer el treball equivalent a dues avaluacions, amb el que això representa pel que fa a recursos, temps i esforç.

En relació amb el temps:

- Ha d'haver **passat prou temps** perquè els impactes s'hagin pogut produir. En termes generals, les intervencions públiques sobre problemes socials requereixen força temps perquè es produeixin canvis observables.
- Alguns impactes es produeixen de forma ràpida però també desapareixen de forma ràpida, cosa que obliga a utilitzar **múltiples mesures en el temps** o a esperar si les dades tornen a la normalitat després d'un increment sobtat.

En relació amb la disponibilitat de dades:

- L'administració o les organitzacions prestadores dels serveis han d'haver recollit les **dades necessàries en arxius** sobre les persones, empreses o

altres unitats individualment considerades. També poden haver recollit les dades per mitjà de qüestionaris d'enquesta.

- Es pot **valorar** si és possible **iniciar un procés de recollida de dades** per mitjà de qüestionaris i enquestes preparats a mida de les necessitats de l'avaluació, però requereix temps i pot tenir un cost elevat.

En una avaluació, el temps de preparació de les dades per a poder-les analitzar pot arribar a ser molt més llarg que no el temps d'anàlisi pròpiament dit. En alguns casos pot arribar al 80% del temps total des que s'inicia el treball. Per a aquesta tasca és molt important poder disposar de persones expertes en la gestió de bases de dades que puguin, a partir de petits programes, filtrar i ordenar les dades originals per a deixar-les a punt per a l'anàlisi.

#### **1.4. Impacte, per a qui?**

En la teoria s'insisteix que l'avaluació d'impacte requereix una intervenció estable i continuada i relativament homogènia. Però, en la pràctica, la intervenció per a cada cas individual, tant si són persones, empreses com barris, pot ser força diferent.

Per exemple, si analitzem els aturats no tots són iguals: n'hi ha que són persones directament ocupables que acaben de perdre la feina; altres són aturats de llarga durada amb greus problemes econòmics, de salut o familiars. Algunes persones són joves però tenen nivells formatius molt baixos. Altres tenen edats força avançades i el pas del temps sense treballar fa que cada vegada els sigui més difícil reenganxar-se perquè necessiten posar-se al dia, ells mateixos senten que estan perdent el tren del treball formal i estable.

Un altre exemple, les empreses presenten grans diferències pel que fa al volum de treballadors i activitat però també respecte a la capacitat d'innovar o d'exportar. Les formes de participació en una política pública que promoció aquests dos aspectes poden ser força diferents. De fet, la política pretén que facin un salt qualitatiu però per a analitzar si es produeix aquest salt és imprescindible agrupar-les correctament i establir-ne categories que tinguin sentit des de la perspectiva de la intervenció pública.

Com que les intervencions públiques poden ser efectives amb determinats tipus de beneficiaris i no amb d'altres, l'avaluació d'impacte comporta analitzar no solament si el programa funciona, sinó **per a qui funciona** mitjançant la desagregació de l'estimació de l'impacte per a diferents grups de població.

## 2. Com es fa?

### 2.1. Tres grans aproximacions a l'avaluació d'impacte

L'avaluació d'impacte pròpiament dita requereix mètodes d'investigació quantitativa. Això no exclou l'oportunitat d'utilitzar mètodes qualitativa per a altres finalitats o tipus d'avaluació o per a entendre millor la política abans d'avaluar-ne l'impacte. D'entrada es pot considerar que hi ha **tres grans aproximacions** metodològiques a l'avaluació d'impacte quantitativa:

- Els experiments reals amb selecció aleatòria prèvia.
- Els quasiexperiments que es basen en les comparacions abans-després i en els grups de control però sense selecció aleatòria prèvia.
- L'anàlisi multivariant per mètodes estadístics en què de fet no hi ha comparació de grups diferents sinó control estadístic de les dades referides a un únic grup.

Les avaluacions per comparació de grups tenen un gran avantatge sobre les avaluacions que només tenen dades per a un sol grup i intenten mesurar els impactes per mitjà de controls estadístics ja que, com que tenen un referent amb el qual comparar, aconsegueixen reduir les amenaces a la validesa dels resultats i per tant a la possibilitat d'equivocar-se en inferir una relació causal entre una intervenció pública i uns impactes observats.

Com sol passar en recerca aplicada, hi ha una **relació inversa entre el que és factible i el que donaria uns millors resultats en termes de validesa**.

- L'experiment real és el millor des de la perspectiva del control de les amenaces (explicacions alternatives) de les conclusions, però és molt difícil d'organitzar en la pràctica ja que és necessari agafar tota la població potencialment beneficiària de la política o programa i fer dependre'n el fet que un individu rebi o no rebi la prestació d'un procediment d'assignació aleatòria. Una vegada dut a terme el procés de selecció dels "afortunats" que reben el programa, aquells als quals no toca rebre la prestació formen el grup de control i fan la funció de contrafactual.
- Els quasiexperiments controlen un menor nombre d'amenaces o explicacions alternatives pel fet que rebre o no la prestació no depèn del procediment aleatori. En aquest sentit són pitjors que els experiments reals. Però els quasiexperiments són millors que l'avaluació per mitjà de mers procediments o controls estadístics ja que es basen en la construcció d'un grup

de comparació que malgrat no ser perfectament equivalent, compleix la funció de contrafactual, una funció que compleixen millor en la mesura que els individus que es comparin siguin més similars.

- El més practicat no deixa de ser l'anàlisi estadística de tipus de regressió múltiple de les variables referides a un únic grup. Aquesta aproximació és més factible però molt menys potent ja que en lloc d'intentar veure què hauria passat en absència de la intervenció, el que es fa és veure l'associació entre determinades característiques dels individus o dels programes que els són aplicats i una o diverses variables que representen l'èxit buscat. La gran diferència rau, però, en el fet que l'associació no és el mateix que la causalitat. Dues coses poden variar al mateix temps i fins i tot en el mateix sentit però això no vol dir que una causi la variació en l'altra, ja que és perfectament possible que sigui un tercer factor el que les faci variar totes dues.

## 2.2. Experiments reals amb selecció aleatòria prèvia

Com s'ha dit, fer una avaluació amb un disseny experimental requereix que els avaluadors controlin el programa i que els gestors, administradors o professionals permetin que s'organitzi un veritable experiment amb els seus usuaris. La idea bàsica és que abans que comenci el programa es posi en marxa un procediment aleatori pel qual es decideixi, entre tota la població potencialment beneficiària, quines persones reben la prestació i quines no la reben.

El fet de recórrer al procediment aleatori és el que fa que es puguin considerar equivalents els individus del grup de tractament i del grup de control en totes les variables que podrien influir en els impactes d'interès perquè si, passat el temps, es detecten diferències es puguin imputar a l'única cosa que diferencia els dos grups i que no és altra que haver rebut la prestació analitzada. La distribució aleatòria de les persones entre els dos grups permet no haver de preocupar-se ni dels factors psicològics difícilment controlables com la motivació, ja que les persones més motivades i les persones menys motivades també es distribueixen de manera similar entre els dos grups. Com se sol dir en la literatura, les amenaces a la validesa interna s'eliminen per la manera com es "construeix" la comparació. Les amenaces a la validesa externa, en canvi, no són controlades pel procediment aleatori. Que els resultats de l'experiment es puguin generalitzar depèn del fet que les persones que participen en l'experiment siguin representatives del total de la població que té aquell problema.

Per exemple, Joan Carles March i el seu equip de l'Escola Andalus de Salut Pública va aconseguir el permís per administrar de forma controlada heroïna en lloc de metadona a persones de la zona de Granada que feia molts anys que hi eren addictes.

Però per a poder fer aquest experiment, l'equip recercador va haver de fer el següent:

- Seleccionar individus susceptibles de beneficiar-se del programa d'administració controlada d'heroïna entre persones que havien fracassat com a mínim dues vegades amb tractaments alternatius.
- Assignar-los amb el seu consentiment, mitjançant un procediment aleatori, a un de dos grups: el grup de tractament en què els subjectes van rebre de forma controlada, per personal sanitari del servei de salut, la dosi d'heroïna tots els dies de l'any; i el grup de control, on els individus rebien el tractament estàndard amb metadona.
- Deixar passar el temps i comparar en un grup i l'altre els impactes d'interès que en aquest cas tenien a veure amb la qualitat de vida, acceptació per la família i reinserció laboral, l'estat de salut i els ingressos a urgències o la implicació en activitats delictives i en processos judicials.

Com es pot veure, la lògica seguida és molt semblant a la dels assajos clínics. De fet, l'exemple és de salut però existeixen experiments reals en altres àmbits com l'educació, la lluita contra la pobresa o la formació ocupacional.

Els experiments reals presenten moltes dificultats de tipus pràctic ja que no és fàcil mantenir unes condicions de "laboratori" quan es tracta amb persones reals que tenen un problema. La principal dificultat deriva del fet que aquestes persones tenen mobilitat, poden canviar de lloc de residència, desaparèixer o simplement no col·laborar.

A més dels problemes pràctics, sovint s'assenyalen els problemes legals i ètics. En alguns països està explícitament prohibit experimentar amb persones reals en alguns àmbits i és possible que es consideri inadequat utilitzar un mecanisme aleatori per a decidir si una persona ha de rebre o no una prestació. En realitat, però, el problema ètic no es planteja tan sovint ja sigui perquè el grup de control continua rebent la prestació que es donava (en l'exemple, la metadona) o perquè, per manca de recursos, no és possible atendre d'una sola vegada tota la població i és necessari fer-ho gradualment.

### 2.3. Quasiexperiments

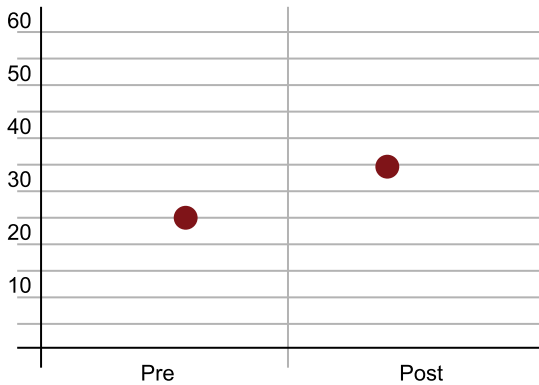
Quan no és possible organitzar un experiment real amb aleatorització prèvia es pot intentar controlar algunes amenaces a la validesa interna per mitjà de dissenys quasiexperimentals. L'estratègia bàsica en aquesta cas és o bé fer una comparació abans-després per a un mateix grup que és objecte d'una d'intervenció, o bé imitar l'experiment real construint *a posteriori* un grup de control tan similar com sigui possible al que és objecte de la prestació, però sense assignació aleatòria prèvia i fer una comparació. Lògicament, també es poden combinar les dues estratègies i fer comparacions abans-després amb grups de comparació construïts *a posteriori*.

La **comparació abans-després** requereix disposar d'informació relativa als impactes de la intervenció abans que es porti a terme i passat un temps després que s'implementi.

Per exemple, en el cas dels programes educatius, normalment existeix informació sobre rendiment escolar (notes dels alumnes en lectura, matemàtiques, etc.) prèvia a la implementació d'una reforma i la mateixa informació per a un moment posterior.

Un altre exemple, si ho permet el pressupost es pot fer una enquesta preintervenció i una enquesta postintervenció per tal de generar la línia base (*baseline*) i tenir un referent per a comparar posteriorment.

Gràfic 1. Comparació abans-després



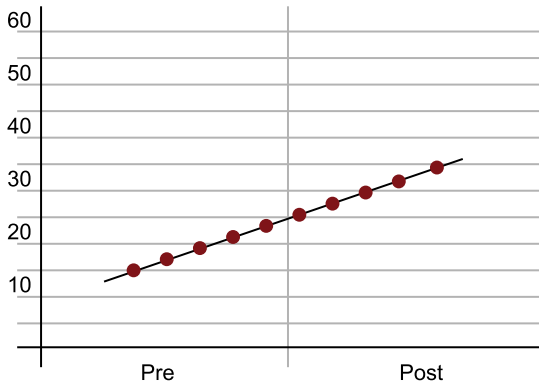
Font: elaboració pròpia

El principal problema del disseny que es limita a fer una comparació abans-després –amb dues dades en dos moments puntuals– és que no permet controlar diverses amenaces a la validesa interna i qualsevol inferència que es faci pot ser qüestionada.

Per exemple, la Generalitat va incentivar la creació de brigades de voluntaris per apagar incendis forestals durant els anys noranta. Si tenim informació sobre el nombre d'hectàrees cremades per a l'any anterior a la creació de les brigades i informació per a un any posterior, podria passar que s'hagi cremat menys hectàrees però difícilment es podria atribuir al programa, ja que és evident que altres factors (pluviometria, temperatura, humitat, presència de vent, etc.) poden haver influït en el resultat observat.

Una millora sobre la mera comparació entre dos punts, un abans i un després, és la **comparació de múltiples punts abans i després**. De fet, es tracta d'un disseny molt similar a l'anterior però com que es disposa de més punts permet veure'n l'evolució o tendència, és a dir, el comportament normal sense intervenció. La qüestió rau aquí en el fet que es produeixi una fractura o canvi de tendència a partir del moment en què es produeix la intervenció, cosa que permet començar a pensar que sigui el programa el responsable del canvi.

Gràfic 2. Sèrie temporal per a un únic grup

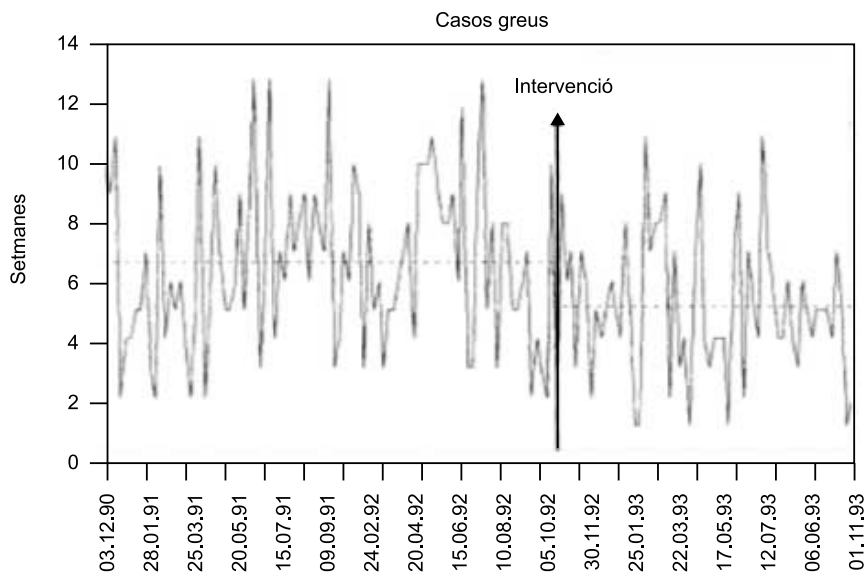


Font: elaboració pròpia

Per exemple, l'autor va avaluar l'eficàcia de la llei que va obligar l'any 1992 tots els motociclistes a portar casc en totes les vies urbanes i interurbanes (Ballart i Riba, 1995). Per a fer aquesta avaluació es va utilitzar dades de la policia sobre les conseqüències dels accidents en què havien participat motociclistes i que aleshores es classificaven en cinc categories: morts, molt greus, greus, menys greus i lleus.

Un primer gràfic (*plot*) de les dades al llarg de més de quatre anys no mostrava cap tendència regular però si es comparava el període abans i després de la prohibició de circular sense casc, s'observava que el nombre de casos greus, molt greus i morts sumats havia baixat, que el núvol de punts es movia en un nivell més baix.

Gràfic 3. Accidents greus motocicletes menys de 125 cc

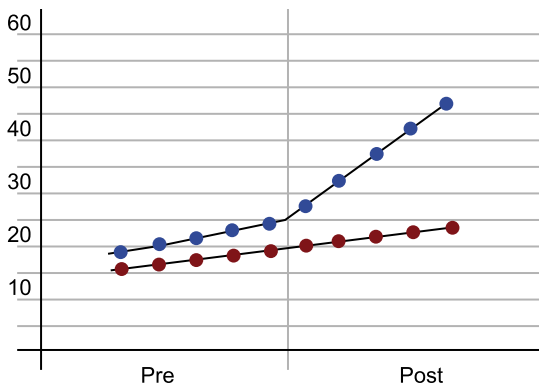


Font: Ballart i Riba (1995)

Malgrat tenir una sèrie temporal de dades, en aquest cas caldria tenir en compte les mateixes amenaces a la validesa i particularment la història, ja que podrien ser factors externs (menys pluja, millora de l'asfalt, millora tècnica de les motocicletes, més atenció per part de la policia) el que estigués al darrere de la baixada observada per les conseqüències de mort o greus.

Per a controlar els factors externs, la clau rau a **construir un grup de comparació** que s'assembli tant com sigui possible al grup que rep la prestació. Això no és possible quan la política és universal, però sí que ho és quan no tothom ha estat objecte de la intervenció i es disposa d'informació tant sobre individus que han estat objecte de la política com d'individus que no ho han estat.

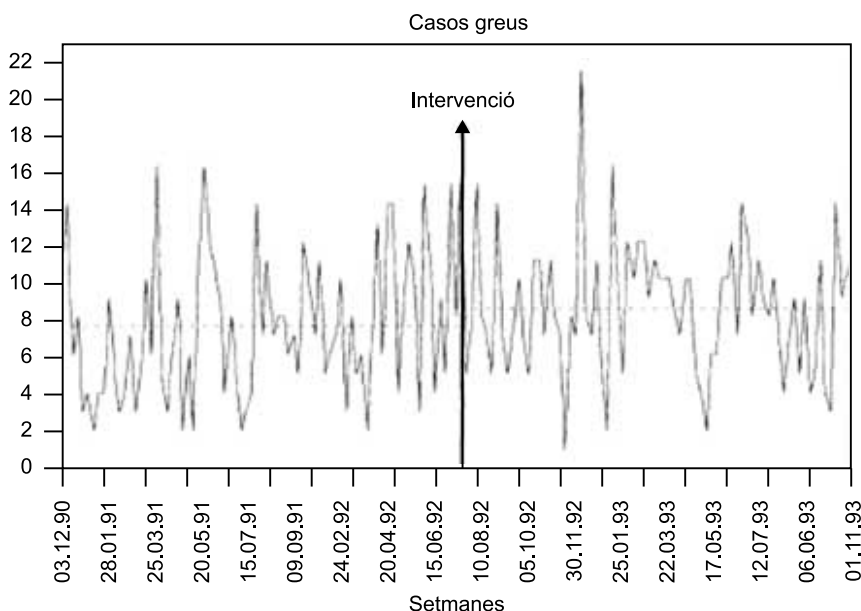
Gràfic 4. Sèrie temporal amb grup de comparació



Font: elaboració pròpia

En l'exemple del cas, la llei només es dirigia als conductors de motocicletes i ciclomotors de menys de 125 cc als quals obligava a portar casc en ciutat ja que la resta de conductors de cilindrades més grans ja hi estaven obligats en tots els casos. Va ser possible, doncs, comparar les sèries temporals durant més de quatre anys dels casos en què l'informe policial indicava conseqüències greus (morts, molt greus i greus) i dels accidents d'aquests dos grups de motociclistes i observar que, mentre per al grup de les motocicletes petites hi havia una fractura en la sèrie de dades en el sentit de davallada, per al grup de les motocicletes grans aquest canvi de tendència no es produïa i es mantenia un comportament molt similar abans i després de l'aprovació de la llei o fins i tot amb una petita tendència a pujar.

Gràfic 5. Accidents greus de motocicletes més de 125 cc



Font: Ballart i Riba (1995)

La idea de comparar amb un grup similar és simple però no fàcil de portar a terme, ja que cal tenir en compte les característiques que siguin rellevants per al cas objecte d'estudi. Una vegada està clar que els grups són comparables i estan ben construïts, si es comprova que existeix una diferència significativa

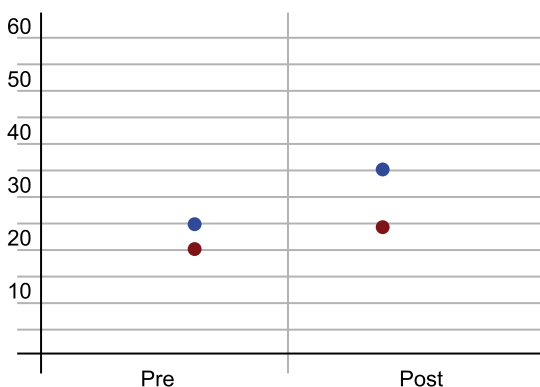


entre els grups que es comparen pel que fa als impactes, en les seves puntuacions abans i després de la intervenció, es pot afirmar que el programa ha estat efectiu.

Així, per exemple, en el cas d'un programa de formació ocupacional per a joves en què els participants són voluntaris, caldria pensar en les característiques que haurien portat aquells individus a participar en el programa (nivell educatiu, motivació, zona de procedència...) o que podrien millorar els seus resultats finals.

La millor manera de construir un grup de comparació *a posteriori* és utilitzant la tècnica del *matching*, que consisteix a buscar per a cada individu d'una mostra dels participants en un programa una parella que s'hi assembli tant com sigui possible. Lògicament, si el *matching* es fa bé (fet que es pot comprovar estadísticament) les parelles seran semblants pel que fa a les variables considerades i per a les quals es tenen dades (per exemple, és difícil tenir dades per a la motivació individual). Els resultats no seran, però, tan bons com en un cas d'assignació aleatòria on el procediment distribueix aleatòriament tots els factors, tant aquells en els quals s'ha pensat com aquells en els quals no s'ha pensat.

Gràfic 6. Comparació abans-després amb grup de comparació



Font: elaboració pròpia

Per exemple, es podria veure l'efectivitat d'haver cursat un màster de gestió pública comparant el tipus de feina i/o el nivell retributiu dels seus graduats en un període de deu anys amb els d'una mostra d'individus que van acabar la mateixa carrera amb un nivell de notes similar però que van decidir no estudiar un màster o estudiar un altre tipus de màster.

Hi ha molts dissenys quasi experimentals que han estat tractats en manuals de referència (en castellà: Alvira Martín, 1991, Ballart, 1992; en anglès: Cook i Campbell, 1979, Khander, Koolwal i Samad, 2010). Els diferents dissenys són el resultat de combinacions diferents pel que fa a les comparacions abans-després, amb un o diversos punts, sobre la base de sèries temporals, tenint en compte les variacions més que el nivell en un punt concret. La idea és descobrir les discontinuïtats, el salt qualitatiu, el canvi de trajectòria i si les diferències observades són significatives.

## 2.4. Anàlisi multivariant

En el cas que no sigui possible comparar grups es pot analitzar per mitjà de models multivariants si, per a un mateix grup d'usuaris o de casos, variacions en la prestació que reben o en les característiques personals tenen un impacte diferencial en els *outcomes* que es persegueixen. En aquest cas es crea un contrafactual per mètodes estadístics ja que de fet s'està comprovant l'associació (no la causalitat) entre dues variables, per exemple trobar feina i haver participat en un programa de pràctiques en una empresa, quan es mantenen constants la resta de variables (edat, sexe, nivell d'estudis, haver rebut cursos d'una administració o d'una ONG).

La forma de procedir consisteix a incloure totes les variables personals i institucionals (del programa) que es considerin rellevants i per a les quals es tinguin dades en un model de regressió. Aquest és una equació en la qual la variable dependent ( $Y$ ) és l'èxit de la política o programa (per exemple, trobar feina) i la resta de variables ( $X_1, X_2, \dots, X_n$ ) indiquen, per cada individu, si ha participat en alguna variant de la intervenció (si l'administració li ha prestat una sèrie de serveis) i les característiques personals que puguin estar associades amb l'èxit (sexe, edat, experiència laboral prèvia...).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon$$

El model de regressió multivariant proporciona els coeficients ( $\beta$ ) per cada una de les variables independents considerades.

Per exemple, el programa busca que els seus participants trobin una feina remunerada la variable *fer pràctiques* té un coeficient positiu de 2,8. Aquest coeficient s'interpreta com que el fet de fer pràctiques incrementa en 2,8 punts la probabilitat de trobar feina quan totes les altres característiques programàtiques o personals no varien. Alhora, es dona aquest resultat amb un cert nivell de significació estadística que no vol dir altra cosa que la probabilitat que aquest resultat no sigui cert i s'hagi donat per pura mala sort.

El principal problema es produeix amb la impossibilitat d'inclusió de les variables rellevants per falta de dades o perquè no són observables. D'altra banda, normalment es disposa de molta informació sobre les característiques personals dels individus (sobre les quals és més difícil intervenir) i en canvi es disposa de poca informació sobre variants de la intervenció que permetin veure quines funcionen millor.

### 3. Lectures obligatòries

#### Presentació

Per aquest mòdul es proposen tres lectures obligatòries. En primer lloc la guia pràctica núm. 5 d'Ivàlua sobre avaluació d'impacte. Com en altres ocasions, aquest text permet reforçar l'explicació del mòdul i introduir l'estudiant en alguns dissenys específics. Com en la resta de mòduls es proposen dues avaluacions reals i força relacionades. Per una banda, la primera avaluació del programa de rendes mínimes de Catalunya sobre la base d'una anàlisi multivariant (publicació de Riba, Ballart i Blasco). Per altra banda, una avaluació dels plans d'ocupació de Catalunya entre 2005 i 2007 realitzada per Ivàlua i dirigida per Blasco.

#### Lectura 1

*Guia pràctica 5. Avaluació d'impacte* (2009). Barcelona: Institut Català d'Avaluació de Polítiques Públiques (Ivàlua).

#### Lectura 2

C. Riba; X. Ballart; Jaume Blasco (2011, gener-març). "Minimum Income and Labour Market Integration Processes: Individual and Institutional Determinants". *Reis* (núm. 133, pàg. 43-58).

#### Lectura 3

*Plans d'ocupació 2005-2007. Informe final d'avaluació* (2010). Barcelona: Institut Català d'Avaluació de Polítiques Públiques (Ivàlua).

#### Preguntes que ha de contestar l'estudiant

- 1) Expliqueu la idea de contrafactual. És possible construir un contrafactual sense comparar grups diferents?
- 2) Quin és el paper de l'anàlisi dels perfils en les dues avaluacions proposades?
- 3) Quina és la principal conclusió que traieu sobre el programa de la renda mínima?
- 4) Descriviu com s'ha de procedir per a dur a terme un procés de *matching* amb la tècnica del *propensity scoring*?



## Bibliografia

**Alvira Martín, F.** (1991). *Metodología de la evaluación de programas*. Madrid: CIS.

**Ballart, X.** (1992). *¿Cómo evaluar programas y servicios públicos? Aproximación sistemática y estudios de caso*. Madrid: Ministerio de Administraciones Públicas.

**Ballart, X.; Riba, C.** (1995). "Impact of legislation requiring moped and motorbikes riders to wear helmets". *Evaluation and Program Planning* (vol. 18, núm. 4, pàg. 311-320).

**Cook, T.; Campbell, D.** (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.

**Khandker, S. R.; Koolwal, G. B.; Samad, H. A.** (2010). *Handbook on impact evaluation*. Washington, DC: The World Bank.

