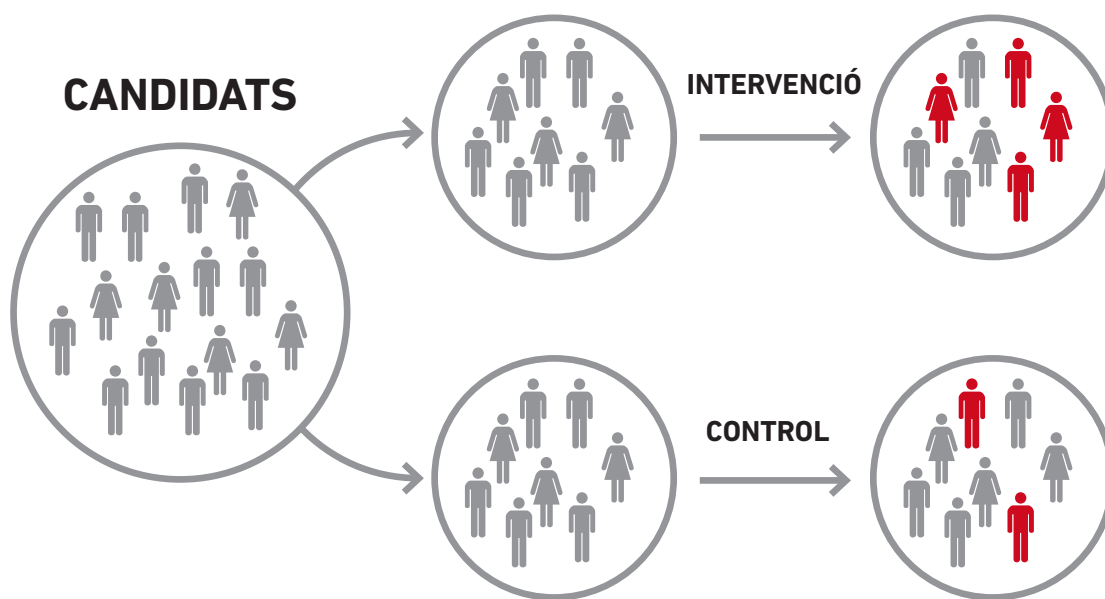


Guia pràctica 10 - Avaluar l'impacte de les polítiques actives d'ocupació

Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques



ivàlua Institut Català d'Avaluació de Polítiques Públiques

SOC

Servei d'Ocupació de Catalunya

Institucions membres d'Ivàlua



© 2013, Ivàlua

No es permet la reproducció total o parcial d'aquest document, ni el seu tractament informàtic, ni la seva transmissió en qualsevol forma o per qualsevol mitjà, ja sigui electrònic, mecànic, per fotocòpia, per registre o altres mètodes, sense el permís del titular del Copyright.

Autors:

David Casado,
Federico Todeschini

Maquetació i disseny portada: jaumbadosa.es

Primera edició: Juny 2013

Amb la col·laboració de:



Dipòsit legal: B-22902-2013

ÍNDEX

1. INTRODUCCIÓ A LA GUIA	PÀG. 5
2. QUÈ VOL DIR AVALUAR L'IMPACTE I COM NO S'HA DE CALCULAR	PÀG. 7
3. DISSENYS ALTERNATIUS PER AVALUAR IMPACTES	PÀG. 11
3.1 AVALUACIONS EXPERIMENTALS	pàg. 11
3.2 REGRESSIÓ DISCONTÍNUA	pàg. 15
3.3 MATCHING	pàg. 17
3.4 MODEL DE DOBLES DIFERÈNCIES	pàg. 18
3.5 ELECCIÓ ENTRE MÈTODES	pàg. 19
4. COM MILLORAR L'AVAUACIÓ DE LES PAO ARA I AQUÍ?	PÀG. 22
4.1 DISSENYAR LA POLÍTICA I L'AVAUACIÓ ALHORA	pàg. 22
4.2 LES EXPERIÈNCIES PILOT: NO MÉS OCASIONS PERDUDES	pàg. 23
4.3 L'EXCÉS DE DEMANDA I L'ALEATORITZACIÓ COM A MECANISME D'ASSIGNACIÓ JUST	pàg. 24
4.4 ELS SUPLENTS: FENT DE LA NECESSITAT VIRTUT	pàg. 24
4.5 LA MILLORA DE LES BASES DE DADES	pàg. 26
4.6 MÉS ENLLÀ DE L'IMPACTE: L'AVAUACIÓ ECONÒMICA DE LES PAO	pàg. 28
5. CONCLUSIONS	PÀG. 30
6. PER SABER-NE MÉS	PÀG. 32
BIBLOGRAFIA	PÀG. 33
ANNEX EXEMPLES D'AVAUACIONS DE POLÍTIQUES ACTIVES D'OCUPACIÓ	PÀG. 35

EXEMPLE 1. DISSENY EXPERIMENTAL 35

EXEMPLE 2. REGRESSIÓ DISCONTÍNUA 37

EXEMPLE 3. MATCHING 39

EXEMPLE 4. MODEL DE DOBLES DIFERÈNCIES 41

1. INTRODUCCIÓ A LA GUIA

En un context com l'actual, amb uns pressupostos públics llastats per la crisi econòmica, els recursos destinats a finançar les polítiques actives d'ocupació (PAO) han experimentat una disminució molt notable, tant a Catalunya com a la resta de l'estat espanyol. Tanmateix, els responsables polítics, independentment del seu color, insisteixen en la necessitat de «millorar l'efectivitat» dels programes o «guanyar en eficiència» en la prestació dels serveis. Aquesta insistència ens sembla positiva, tot i que dubtem que aquest èmfasi es mantingui quan vinguin temps millors, la qual cosa seria un error: preguntar-se si una determinada política activa augmenta la inserció laboral dels aturats (efectivitat) o si, per comparació a d'altres intervencions, el cost per aturat inserit d'aquesta política és més o menys favorable (*eficiència*) són qüestions que l'Administració s'hauria de plantejar en tot moment, sigui quina sigui la situació econòmica i l'estat de les finances públiques. Si no se'n preocupa, correm el risc de seguir finançant programes que no generen els efectes pretesos o, alternativament, deixar de finançar intervencions que els aconsegueixen amb escreix.

Fins al moment, tot i que en termes relatius les PAO han estat més avaluades que d'altres polítiques públiques, com ara l'educació o els serveis socials, estem força lluny dels nivells que s'observen a d'altres països del nostre entorn. Per exemple, en una recent revisió de les avaluacions d'impacte de PAO portades a terme a Europa (Kluve, 2010), en la qual únicament s'inclouien aquelles que superaven cert llindar de rigor metodològic, tant sols hi ha 3 avaluacions espanyoles d'un total de 137. La situació a Catalunya és lleugerament millor, arran de les avaluacions impulsades pel SOC durant el període 2007-2012, realitzades tant per Ivàlua com per d'altres institucions. Però hi ha pocs motius per a l'autocomplaença, ja que la percepció per part del principal finançador de les PAO, la Unió Europea, és que el camí a recórrer per part d'Espanya en aquest sentit és gran; de fet, fins i tot el *Memorandum of Understanding* (MoU) s'ha fet ressò d'aquesta situació, i en un dels seus punts demana a Espanya impulsar avaluacions rigoroses sobre l'efectivitat de les PAO.

En definitiva, no tant sols mesurar l'impacte és quelcom important, sinó que la pressió per a fer-ho, i fer-ho bé, anirà en augment de forma accelerada. En aquest context, el principal propòsit de la guia que teniu a les mans és explicar a una audiència no especialitzada, composta per polítics i gestors, els principals aspectes de l'avaluació d'impacte, posant un èmfasi especial en la seva aplicació a l'àmbit de les PAO¹. En concret, després de descriure què vol dir avaluar l'impacte o efectivitat d'un programa (apartat 2), es descriuen els principals dissenys que poden utilitzar-se a tal efecte (apartat 3). Finalment, es discuteixen els principals elements que, al nostre parer, haurien de tenir-se en compte per tal d'impulsar un esquema estable d'avaluacions d'impacte de les PAO (apartat 4). En aquest sentit, la guia constitueix un complement al document "Model d'avaluacions del SOC", el qual també ha estat elaborat per Ivàlua recentment².

Notes:

¹ Els lectors interessats en aprofundir en els aspectes més tècnics de l'avaluació d'impacte trobaran al final de la guia una selecció comentada de lectures.

² La guia està centrada en l'avaluació d'impacte o d'efectivitat. Tanmateix, cal tenir present que existeixen altres tipus d'avaluacions, com ara l'avaluació de necessitats, la d'implementació, l'avaluació ex-ante o l'avaluació econòmica, que exploren altres qüestions igualment importants a l'hora de valorar una política pública. Les guies 2, 4, 6 i 7 de la "Col·lecció Ivàlua de guies pràctiques sobre avaluació de polítiques públiques" ofereixen, respectivament, sengles introduccions al respecte. Aquestes guies es troben disponibles a <http://is.gd/qWVvkR>.

2. QUÈ VOL DIR AVALUAR L'IMPACTE I COM NO S'HA DE CALCULAR

La majoria de polítiques públiques, i les PAO no en són una excepció, pretenen en darrera instància modificar algun *outcome* que es considera rellevant. Per exemple, els programes del SOC que tenen com a destinataris els joves desocupats busquen, mitjançant actuacions diverses, millorar el seu grau d'inserció laboral. En aquest sentit, si per exemple existís un hipotètic programa anomenat Segona Oportunitat, que consistís en proporcionar 4 mesos de formació professionalitzadora a joves aturats sense la ESO, com valorar si el programa millora o no la inserció laboral?

Una aproximació ben senzilla per mirar de respondre aquesta pregunta consisteix a valorar la situació laboral dels beneficiaris del programa un temps després d'haver-lo acabat, com ara passats 6 mesos. Suposem que en el cas de Segona Oportunitat, per exemple, un 43% dels beneficiaris es trobessin treballant. Podem afirmar que haver assolit aquest nivell d'inserció és atribuïble exclusivament a la participació dels joves en el programa? No, en absolut. El fet que els joves beneficiaris hagin trobat o no una feina pot dependre de moltes coses que no tenen res a veure amb el programa, com ara la situació del mercat laboral, l'existència d'una política estatal que incentivi la contractació juvenil, o el nivell d'intensitat en la recerca de feina, entre d'altres factors. De fet, un cop es reconeix aquesta multiplicitat de causes, la qüestió rellevant és si la taxa d'inserció laboral dels joves de Segona Oportunitat hagués estat o no del 43% en cas de no haver participat en el programa: només en el cas que aquesta fos inferior, podríem afirmar realment que el programa ha tingut un *impacte* positiu sobre la inserció laboral dels beneficiaris de Segona Oportunitat; de fet, si la taxa d'inserció en absència del programa fos superior al 43%, l'*impacte* seria negatiu.

En termes més generals, l'*impacte* d'una intervenció o programa és la diferència entre allò que realment esdevé als participants i l'anomenat *contrafactual*: és a dir, allò que els hagués succeït si no haguessin participat. Tanmateix, atès que és impossible que els mateixos subjectes participin i no participin en un determinat programa, l'*outcome* contrafactual no és observable. El principal repte de l'avaluació d'*impacte*, per tant, consisteix a proposar una estratègia que permeti obtenir una mesura creïble d'aquest *outcome* contrafactual; tot seguit, si de l'*outcome* observat es resta aquesta estimació de l'*outcome* contrafactual, s'obté una mesura de l'*impacte* real del programa.

En aquest sentit, tal com ens ha permès il·lustrar l'exemple de Segona Oportunitat, sembla clar que utilitzar les taxes d'inserció laboral post-programa com a mesura de l'*impacte* d'una PAO resulta completament erroni, excepte si estem disposats a defensar que l'únic factor que determina la inserció laboral dels aturats siguin aquestes polítiques (o, amb d'altres paraules però igual sentit, que cap dels beneficiaris del programa hagués trobat feina si no hi hagués participat). Tanmateix, la llista de procediments incorrectes per mesurar l'*impacte* d'una PAO no s'esgota en l'ús de la inserció laboral post-programa. Hi ha dos procediments addicionals

que, malgrat reconèixer la importància de definir un contrafactual, condueixen a estimacions errònies d'impacte.

Una primera aproximació consisteix a preguntar directament als beneficiaris, un cop la seva participació en el programa ha finalitzat, si fruit d'aquesta participació la seva ocupabilitat ha millorat o no. En aquest sentit, tot i que en el nostre context no es realitzen enquestes d'aquests tipus, a d'altres països, com ara els EUA, constitueixen un instrument rutinari de seguiment de la majoria de PAO (Smith, 2004). En el fons, s'espera dels individus que estimin l'impacte de llur participació en el programa, a partir de fer més o menys explícita la idea de *d'outcome* contrafactual. Tanmateix, malgrat a priori la idea de preguntar directament als beneficiaris podria resultar atractiva, l'evidència disponible sobre la robustesa dels resultats d'aquestes auto-avaluacions d'impacte resulta decebedora: així, quan es comparen els resultats obtinguts mitjançant enquestes amb els impactes estimats per als mateixos beneficiaris mitjançant avaluacions experimentals, que com veurem més endavant constitueixen la metodologia més sòlida per a estimar impactes, la conclusió és que els individus no són (som) gaire bons avaluant impactes (Heckman i Smith, 1998). Aquest resultat no ens hauria de sorprendre massa: d'una banda, els individus tenen tota mena de problemes cognitius a l'hora de prendre bones decisions sobre qüestions força més senzilles que estimar contrafactuals, com palesen les nombroses investigacions portades pels psicòlegs i economistes del comportament; d'altra banda, quan els beneficiaris d'un programa sospiten que les seves respostes poden condicionar-ne la seva continuïtat, tenen incentius a comportar-se estratègicament i sobre-estimar llurs valoracions dels impactes assolits pel programa.

La segona alternativa que sovint es proposa per mesurar l'impacte d'un programa passa per comparar la inserció laboral dels participants amb un conjunt de no participants, similars en algunes característiques rellevants, com ara l'edat o el seu nivell formatiu. En el cas de Segona Oportunitat, per exemple, podríem comparar la inserció laboral a 6 mesos dels beneficiaris del programa (43%) amb la inserció corresponent de la resta d'aturats de Catalunya, igualment joves i sense la ESO (30%). Tanmateix, constitueix aquest 30% d'inserció laboral un contrafactual creïble de *l'outcome* laboral dels participants o, en d'altres paraules, l'impacte del programa és de 13 punts percentuals?

De nou, la resposta és no, i el motiu té a veure amb el **mecanisme de selecció de beneficiaris** que empen les PAO, entre les quals Segona Oportunitat no tindria per què ser-ne una excepció. En particular, malgrat el nostre hipotètic programa hauria estipulat els criteris bàsics a reunir pels beneficiaris, com ara l'edat, el temps a l'atur i la baixa formació, hi ha dues "forces" que governen el procés de participació en el programa, les quals van molt més enllà dels atributs que haurien aparegut a la convocatòria de Segona Oportunitat.

D'una banda, d'entre tots els joves aturats sense ESO que podrien participar en aquest programa, que com la majoria de PAO també seria de caràcter voluntari, només alguns

d'ells acabarien postulant- s'hi i, en cas de ser escollits, també només alguns dels qui s'hi haguessin postulat hi acabarien participant. La qüestió clau és que tant la decisió de postular-se inicialment com la d'acabar volent participar si són acceptats estan fortament influïdes per factors, com ara la motivació, la renda familiar o la implicació dels pares, que poden ser molt diferents del conjunt de joves aturats no participants, i que de ben segur afecten les probabilitats d'inserció laboral posterior dels individus.

D'altra banda, a més del comportament dels potencials beneficiaris, la selecció de participants també és el reflex de les decisions que es prenen des de la banda de l'oferta, típicament per part dels gestors de les Oficines de Treball o els tècnics de les entitats proveïdores de serveis; de nou, a més dels criteris formals de cada convocatòria, aquests professionals seleccionen d'entre el conjunt d'aturats "elegibles" no tant sols a qui s'informa sobre un determinat programa, sinó també els candidats més adients d'entre aquells que s'acaben postulant, generalment a través de la realització d'entrevistes, atès que típicament hi ha excessos de demanda considerables. La qüestió important és que en aquest procés de selecció, de nou, hi ha múltiples atributs dels candidats que es tindran en compte, com ara la motivació, la bona presència, etc., que també tenen una influència clara sobre les possibilitats de trobar posteriorment una feina.

Així doncs, fruit de la concurrència dels dos processos anteriors, els joves aturats amb baixa formació que acabin participant a Segona Oportunitat poden ser molt diferents de la resta d'aturats que, malgrat ser joves i no tenir la ESO, no hi acabaran participant. I, atès que aquestes diferències influeixen sobre les possibilitats de trobar una feina, la comparació entre la taxa d'inserció laboral dels participants amb la dels no participants no pot considerar-se una mesura vàlida de l'impacte del programa, ja que una part indeterminada de les diferències entre ambdues taxes d'inserció (43% vs 30%) tindrà el seu origen en la diferent composició d'ambdós grups de joves, un fenomen que es coneix amb el nom de "biaix de selecció". Tal i com s'explica en el proper apartat, els mètodes que s'utilitzen per mirar d'avaluar l'impacte d'una política intenten, mitjançant estratègies diverses, definir grups de comparació que no estiguin exposats a aquest "biaix de selecció" que acabem d'esmentar.

Tanmateix, abans de presentar aquest conjunt de dissenys avaluatius, convé deixar ben clar que la pretensió de mesurar correctament l'impacte d'un programa no és una simple dèria acadèmica, pròpia d'investigadors cegats pel rigor científic. El fet de mesurar incorrectament l'impacte d'un programa pot resultar molt contraproductiu per als beneficiaris de les PAO, que en darrera instància haurien de constituir la preocupació fonamental d'acadèmics, polítics i gestors. Suposem, per exemple, que les decisions sobre assignació de recursos entre PAO, o entre proveïdors d'una determinada PAO, es portessin a terme d'acord als nivells d'inserció laboral post-programa. En ambdós casos, a menys que les poblacions ateses siguin completament equivalents, una part de les diferències entre els nivells dels *outcomes* post-programa seran el resultat de diferències en les característiques dels individus (biaix de

selecció) o a l'existència de factors contextuais diferents. De fet, seria perfectament possible trobar situacions en què nivells elevats d'inserció laboral post-programa coexistissin amb impactes nuls o negatius i, alternativament, casos en els quals els nivells d'inserció laboral fossin baixos però els impactes positius. Així doncs, si volem acabar destinant més recursos a les polítiques o als proveïdors que realment "fan més" per incrementar la ocupabilitat dels aturats, resulta fonamental mesurar correctament els impactes. A la propera secció presentem, tal i com havíem avançat, les principals tècniques que l'avaluació d'impacte utilitza per mirar de mesurar correctament l'efectivitat dels programes.

3. DISSENYES ALTERNATIUS PER AVALUAR IMPACTES

El gran repte de l'avaluació d'impacte, com ja ha estat assenyalat, rau en l'estimació d'un contrafactual creïble: això és, una mesura vàlida del què els hi hagués passat als beneficiaris de la política si no hi haguessin participat. En aquest sentit, les PAO constitueixen un camp d'intervenció pública propici per a l'avaluació d'impacte ja que, donat que la majoria de programes són voluntaris i la cobertura és limitada, acostuma a ser possible disposar d'un grup potencial d'individus no participants susceptibles d'actuar com a grup de comparació. D'altra banda, ates que les experiències pilot són habituals en aquest camp, talment que alguns dels nous programes es proven primer en un conjunt limitat de territoris, sorgeixen tot sovint oportunitats de trobar territoris o persones que no han participat en la política que estem interessats a avaluar. De fet, com s'explica als apartats que segueixen, allò que diferencia a les diverses tècniques d'avaluació d'impacte és, bàsicament, la manera com cadascuna d'elles defineix el grup de comparació que s'empra per estimar el contrafactual³. L'exposició de les diverses tècniques es complementa amb sengles exemples il·lustratius, corresponents a avaluacions reals de PAO, recollits i comentats a l'annex d'aquesta guia.

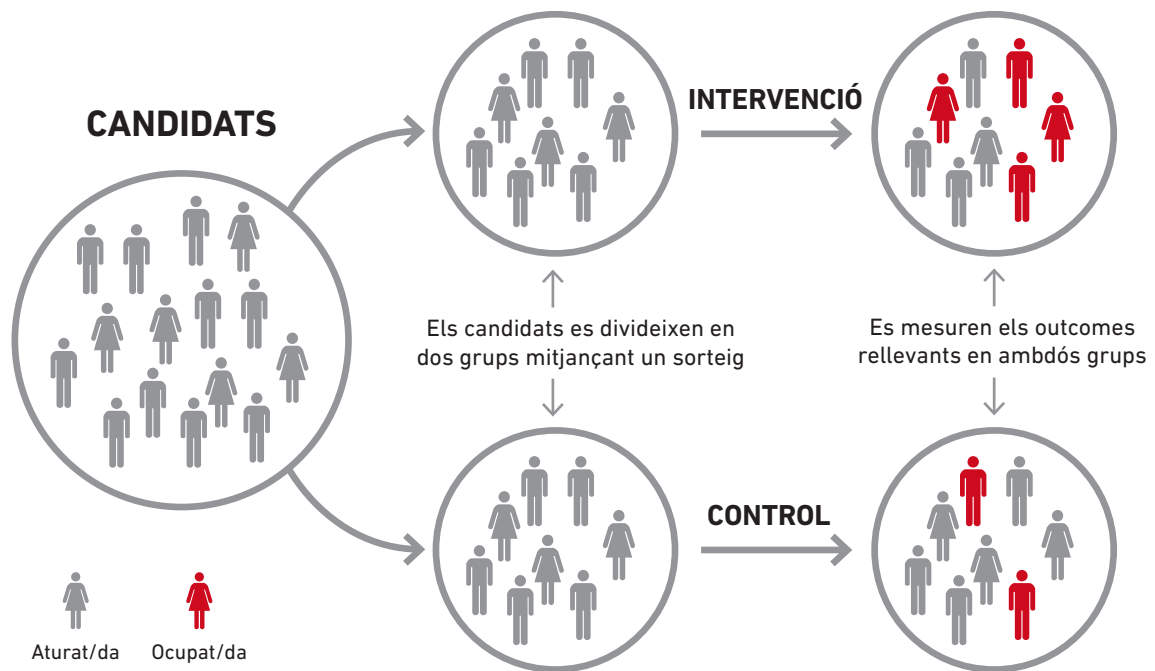
3.1 AVALUACIONS EXPERIMENTALS ⁴

Què és una avaluació experimental?

Avaluar l'impacte d'una política pública mitjançant un disseny experimental és, des d'una perspectiva metodològica, molt similar a aplicar la lògica que segueixen els assajos clínics per testar l'efectivitat d'un fàrmac. La diferència rau en que allò que avaluem no és un fàrmac sinó un determinat programa, i que l'*outcome* sobre el qual busquem produir un efecte no és l'estat de salut sinó alguna problemàtica que n'afecta els destinataris.

Un exemple pot resultar útil per il·lustrar el funcionament d'una avaluació experimental. Així doncs, seguint amb Segona Oportunitat, l'avaluació experimental d'aquest hipotètic programa podria fer-se de la següent manera: 1) donar instruccions a les oficines de treball perquè identifiquessin potencials beneficiaris del programa, amb l'objectiu d'assolir una xifra de 1.000 candidats; 2) mitjançant un procediment aleatori, i previ consentiment dels candidats, aleatoritzaríem la participació en Segona Oportunitat: 500 joves el rebrien i 500 no; i 3) transcorregut un cert temps després de la finalització del programa, comparariem els *outcomes* rellevants, com el grau d'inserció laboral, entre el grup de tractament i de control. La figura 1 mostra gràficament l'essència d'una avaluació experimental del programa Segona Oportunitat.

Figura 1.



Font: Adaptat de Haynes et al. (2012).

Ara bé, per què l'aleatorització, ja sigui de pacients en un assaig clínic o de joves aturats en el nostre exemple, permet avaluar l'impacte d'un tractament o programa d'una forma més vàlida que altres tècniques?

Gràcies a l'aleatorització, un experiment aconsegueix que el grup de tractament i el de control estiguin «equilibrats» en tots aquells atributs personals que poden influir sobre l'*outcome* d'interès, com poden ser, en el cas de Segona Oportunitat, la motivació, l'experiència laboral prèvia o el fet de ser o no immigrant. D'aquesta manera, quan una vegada finalitzat el programa comparem els *outcomes* entre tots dos grups per inferir l'impacte, podem descartar que el resultat obtingut sigui la conseqüència de que ambdós grups són diferents; en d'altres paraules, l'aleatorització ens ha permès eliminar el biaix de selecció que esmentàvem a l'apartat previ. D'altra banda, atès que els dos grups estan exposats als mateixos «factors de context» mentre dura el programa, com podria ser per exemple una millora del mercat laboral en el cas de Segona Oportunitat, també podem descartar que aquests factors siguin els responsables de les diferències post tractament en els *outcomes*. En resum, si detectem aquestes diferències en els *outcomes* entre ambdós grups, els podrem atribuir a l'únic tret que els diferencia: és a dir, haver participat o no en el programa. L'experiment social ens proporciona, per tant, una estimació vàlida de l'impacte del programa.

Què s'entén exactament per aleatorització?

L'aleatorització de la participació constitueix la pedra angular d'un experiment social (ES) i, per a que una avaluació pugui ser considerada experimental, la seva existència és un requisit indispensable. L'aleatorització que caracteritza un ES no s'ha de confondre amb el mostreig aleatori que s'ha d'exigir a una enquesta, ja sigui de salut o de població activa, per a que els resultats obtinguts siguin representatius de la població. D'una banda, mentre que el que ha de ser aleatori en una enquesta és la selecció dels subjectes a entrevistar, en un experiment social s'exigeix que, entre els candidats a participar en el programa, l'elecció d'aquells qui hi acaben participant i d'aquells qui no es dugui a terme mitjançant un procediment aleatori. D'altra banda, l'aleatorització en un ES no busca la representativitat dels resultats, sinó permetre estimar sense biaixos l'impacte del programa avaluat, com ja hem comentat.

Això no vol dir, tanmateix, que un experiment social no pugui aleatoritzar també el procés de captació de candidats. Per exemple, si les 1.000 escoles de Catalunya estiguessin disposades a participar en un programa d'incentius a professors, però només hi hagués pressupost per aplicar l'esquema en 100 centres, podríem escollir 200 escoles a l'atzar d'entre les 1.000 (mostra de candidats) i, a continuació, assignar aleatòriament la participació en el programa a la meitat d'elles. La primera aleatorització conferiria «representativitat» als nostres resultats, en el sentit que es podrien considerar extrapolables a les 800 escoles «no experimentals», però és la segona aleatorització la que ens permet mesurar l'impacte del programa i l'única necessària per definir una avaluació com a experimental.

L'avaluació experimental de les PAO

En el món anglosaxó, especialment als EUA, l'avaluació experimental de polítiques públiques constitueix una pràctica freqüent des de fa més de 30 anys i, de fet, els programes laborals són un dels àmbits més actius al respecte, amb desenes d'intervencions avaluades experimentalment. A Europa, en canvi, les avaluacions experimentals de PAO eren una raresa fins fa poc més d'una dècada. Tanmateix, durant els darrers anys s'han començat a realitzar a Europa nombroses avaluacions experimentals en l'àmbit de les PAO, com l'*Employment Retention and Advancement (ERA) demonstration* duta a terme al Regne Unit (Hendra et al. 2011) o l'avaluació de diversos esquemes d'activació d'aturats realitzats a Dinamarca (Graversen i Van Ours, 2008), per posar només dos exemples. Una menció especial mereix el *Fonds d'Expérimentation pour la Jeunesse*, el qual s'inicià a França l'any 2008 i que, a través d'una dotació de més de 200 milions d'euros per al període 2009-2014, intenta afavorir l'avaluació experimental de noves formes d'intervenció destinades a combatre el fracàs escolar i l'exclusió laboral dels joves francesos. Els programes avaluats fins ara, o en curs d'avaluació, són molt variats i inclouen des d'intervencions orientades a incrementar la implicació dels pares en el procés educatiu fins a la provisió de tutories per prevenir l'abandonament escolar, passant per la prestació de serveis d'orientació laboral a joves aturats. Val a dir que, fins on arriba el nostre

coneixement, encara està per néixer la primera avaluació experimental d'una PAO a l'estat espanyol.

Crítiques als experiments

Un primer aspecte que se sol criticar dels experiments és que són cars. Sens dubte, aquesta apreciació es deu en part al pòsit que van deixar els primers ES duts a terme als EUA, els quals van implicar equips molt nombrosos, realització d'enquestes molt costoses, etc. Avui dia, la informatització dels registres administratius ha permès, en molts casos, esquivar la necessitat de realitzar costoses enquestes sense que això suposi perdre riquesa analítica, ja que l'exhaustivitat i fiabilitat d'aquests registres resulta molt notable. En aquest sentit, cal esmentar que les avaluacions experimentals de PAO europees, com és el cas de les dues esmentades anteriorment, fan servir únicament registres administratius.

En qualsevol cas, més enllà de les consideracions econòmiques, una de les crítiques més habituals envers les avaluacions experimentals té un rerefons ètic: resulta inadequat privar determinats individus (els del grup de control) dels beneficis que suposa una nova política utilitzant un mecanisme tan arbitrari com l'aleatorització. La rèplica per part d'aquells qui veuen en els experiments socials una eina adequada d'avaluació se sustenta en tres consideracions.

La primera és que la presumpció que s'està privant alguns individus d'alguna cosa beneficiosa no hauria de tenir sentit si l'experiment està justificat, ja que és precisament l'absència de dades sobre l'efectivitat del programa el que justifica la seva avaluació. D'altra banda, són poques les ocasions en què pertànyer al grup de control implica no rebre cap tipus d'intervenció, sinó que més aviat el que es compara és la nova política respecte de «seguir com fins ara». Finalment, hi ha situacions força freqüents en què es pot considerar l'aleatorització com un criteri d'assignació equitatiu, com per exemple quan la manca de recursos no permet atendre d'una sola vegada tota la població potencialment beneficiària de la política; de fet, quan es produeixen situacions d'aquest estil, un disseny experimental més acceptable que utilitzar una simple loteria entre individus és optar per un desplegament gradual aleatoritzat: s'aleatoritza el moment del temps en què diferents grups d'individus o territoris començaran a rebre el nou programa.

Una última objecció que habitualment es fa als experiments és que acostumen a no tenir validesa externa o, en paraules menys tècniques, que els resultats que s'obtenen pel que fa a l'impacte d'una política, tot i ser vàlids respecte dels subjectes, moment i lloc en què es va dur a terme l'experiment, poden no ser extrapolables a contextos diferents. Aquells que es dediquen a l'experimentació social han intentat mitigar la manca de validesa externa per dues vies. En primer lloc, són habituals les anomenades avaluacions multi-site, en les quals el programa s'avalua aplicant-lo a llocs diversos (per exemple, diferents municipis), amb l'objectiu

d'analitzar fins a quin punt els resultats d'impacte varien en funció dels contextos. D'altra banda, quan el nombre de rèpliques experimentals d'un determinat tipus de programa és suficientment important, es pot dur a terme el que s'anomena una metanàlisi dels resultats obtinguts, és a dir, un exercici quantitatiu de síntesi que pretén establir si el programa resulta efectiu amb caràcter general, amb independència de les poblacions, llocs i moments en què s'aplica.

Al marge de les crítiques anteriors, és evident que les avaluacions experimentals plantegen reptes importants tant des d'una vessant metodològica com des d'un punt de vista de viabilitat política. Tanmateix, tal i com hem exposat a d'altres llocs (Casado, 2012), on es discuteixen amb detall aquestes altres qüestions, el que sorprèn és que aquest disseny avaluatiu, a bastament utilitzat arreu i considerat el "gold standard" de l'avaluació d'impacte, no hagi estat emprat encara a l'estat espanyol per avaluar cap política pública, incloses les PAO.

La resta de mètodes d'avaluació d'impacte, els quals discutirem tot seguit, comparteixen amb el disseny experimental la definició d'un grup de comparació, integrat per no participants, que també s'empra per estimar l'outcome contrafactual dels participants. L'existència d'un grup de comparació, com ja hem discutit en el cas dels experiments, permet eliminar de l'estimació d'impacte la influència dels anomenats factors contextuais (atur general, altres PAO, etc.), ja que tant els beneficiaris del programa com el grup de comparació hi estan igualment exposats.

En canvi, a diferència del disseny experimental, que elimina el biaix de selecció mitjançant l'aleatorització de la participació, la resta de mètodes només assoleixen aquest objectiu si es compleixen certs supòsits amb relació al mecanisme que governa la participació en el programa. En essència, mentre que les avaluacions experimentals eliminen el risc que suposa el biaix de selecció alterant "per disseny" el procés de captació de participants, la resta de mètodes aspiren a obtenir mesures de l'impacte no contaminades per aquest biaix de la mà de la modelització estadística.

3.2 REGRESSIÓ DISCONTÍNUA

L'anomenada "regressió discontinua" (RD) és probablement la tècnica d'avaluació d'impacte no experimental que ha tingut un creixement més intens en els darrers anys, especialment en l'àmbit educatiu (Schlotter et al., 2010). Aquesta tècnica s'aplica quan la participació en el programa que es pretén avaluar depèn del fet que una variable prengui un conjunt de valors determinats. Per exemple, tornant a l'hipotètic programa Segona Oportunitat comentat prèviament, podria donar-se el cas que els requisits de la convocatòria estipulessin que només els individus amb una renda familiar menor a un cert punt de tall (per exemple, 20.000 euros) hi poden participar. Alternativament, podríem pensar en l'existència d'un barem (per exemple, els individus obtenen punts en funció de la mida de la llar, el nivell educatiu del cap de família, l'edat, etc.) que determina que només poden participar en el programa aquells que obtenen una puntuació superior (inferior) a un cert llindar prefixat a la convocatòria. En qualsevol cas, per

tal de poder implementar la tècnica de la RD, no importa si el punt de tall (de la renda familiar o del barem) és més gran o més petit, sinó que l'única cosa important és que hi hagi un punt de tall.

La lògica del mètode de regressió discontinua és la següent. El fet que la participació sigui determinada per un punt de tall comportarà que hi hagi un salt, justament al llindar, en el percentatge d'individus que participen en el programa. De fet, si el llindar opera de manera estricta, el que observarem al voltant del punt de tall seran dos grups d'individus, uns que participen en el programa i els altres que no. En el nostre exemple, si suposem que el llindar es troba fixat en 20.000 euros, observariem un "salt" en el percentatge de participants just per sobre i per sota d'aquesta quantitat. Doncs bé, si el programa Segona Oportunitat resultés efectiu, aleshores també esperariem que el salt en el percentatge de participació es reflectís en la taxa d'inserció laboral i, per tant, hi hagués també un "salt" entre la taxa d'ocupació dels individus amb una renda familiar just per sobre i per sota del llindar. La idea subjacent de la RD és que la quantitat exacte en la que es fixa el llindar, ja siguin 20.000, 10.000 o 15.000 euros, resulta bàsicament arbitrària. Així doncs, des de la perspectiva dels aturats, el fet de participar o no en el programa acaba depenent de quelcom que "s'assembla a un sorteig", ja que acaben podent accedir o no en funció que la regla fixada per l'administració els deixi per sobre o per sota del llindar. I, si això és cert, cal pensar que aquells que acaben participant i no participant, just als volts del llindar, són essencialment idèntics en tots els atributs rellevants. El biaix de selecció hauria quedat així eliminat.

Tanmateix, com ja hem esmentat prèviament, la tècnica de la RD només proporciona estimacions vàlides si es satisfan un conjunt de supòsits que cal tenir ben presents. En primer lloc, per a que el mètode funcioni és necessari suposar que no hi ha cap altra variable que «salti» en el mateix punt de tall. Per exemple, no podríem aplicar el mètode de regressió discontinua si el punt de tall de 20.000 euros també coincidís amb el punt de tall d'algun altre programa o política (cursos de formació, pensió no contributiva, etc.), ja que si fos així, no podríem saber quina part del «salt» en la taxa d'ocupació és deguda a Segona Oportunitat i quina part a l'altre programa que també mostra un salt. Una altra condició important és que els individus no puguin triar en quin costat del punt de tall es posen. Per exemple, l'estimació quedaria invalidada si els individus més motivats poguessin reduir la renda de 20.050 euros a 19.950 euros amb l'objectiu de poder participar a Segona Oportunitat. En aquest cas, no podríem saber quina part del salt en la taxa d'ocupació és deguda al programa, i quina a la diferent motivació que tenen els individus la renda familiar dels quals és de 20.050 euros comparada amb la dels individus la renda familiar dels quals és de 19.950 euros.

En qualsevol cas, tot i que el mètode de la regressió discontinua proporciona bones estimacions de l'impacte quan es satisfan els supòsits anteriors, cal tenir present que aquesta tècnica tan sols informa sobre l'efectivitat del programa per aquells individus que estan a prop del llindar. Així, seguint amb el nostre exemple, el mètode no proporcionaria estimacions vàlides sobre

l'impacte del programa en el cas d'individus la renda familiar dels quals fos de 30.000 euros o de 10.000 euros, per exemple, atès que estan lluny del punt de tall de 20.000 euros que garanteix la semblança entre participants i no participants sobre la que descansa la bondat del mètode.

Finalment, val a dir que l'aplicació de la RD en l'àmbit de les PAO ha estat menys important, tot i que existeixen algunes excepcions notables, com ara l'avaluació del *New Deal for Young People* (Di Giorgi, 2008) que està resumida a l'annex. La principal raó d'aquest ús limitat cal buscar-la, a parer de Card et al. (2011), en el fet que la participació en la majoria de PAO no acostuma a venir definida per una regla d'elegibilitat estricta, i quan hi és, o bé es tendeix a aplicar de manera laxa, o els candidats la coneixen i miren d'alterar la seva puntuació per poder accedir al programa (manipulant el seu nivell formatiu, de renda, etc.). En qualsevol cas, atès que la RD és una metodologia que permet obtenir estimacions molt robustes d'impacte quan els supòsits en els que es basa es satisfan, val la pena tenir-la present i avaluar-ne la seva aplicabilitat en cada cas particular.

3.3 MATCHING

Aquesta tècnica tracta d'imitar un experiment mitjançant la definició a posteriori, amb l'ajut de l'estadística, d'un grup de comparació de no participants que s'assembli el més possible als beneficiaris del programa que es pretén avaluar. L'aplicació d'aquest mètode pot considerar-se en aquells casos en què, amb posterioritat a l'acabament del programa, disposem d'informació per als participants i els no participants sobre: 1) els *outcomes* amb relació als quals vulguem mesurar l'impacte del programa, com ara la inserció laboral en el cas d'un curs de formació ocupacional; i 2) tots aquells factors (personals, familiars, d'entorn, etc.) que simultàniament hagin pogut influir, d'una banda, sobre el fet d'acabar participant o no en el programa i, de l'altra, sobre els *outcomes* respecte els quals volem avaluar l'impacte (p.ex. inserció laboral).

El que el mètode de *matching* proposa és utilitzar tota la informació anterior per construir un grup de comparació legítim, constituït únicament per no participants que realment s'assemblin als beneficiaris, per poder així estimar l'impacte del programa simplement com la diferència entre els *outcomes* de tots dos grups un cop aquell ja ha acabat. Per fer-ho, el mètode busca, per a cadascun dels participants, una parella o "match" que sigui el més semblant possible en el conjunt de variables (personals, familiars i d'entorn) esmentades prèviament. A tal efecte, poden emprar-se algoritmes estadístics diversos, els detalls dels quals convé deixar al marge a fi i efecte de mantenir el to divulgatiu que hem volgut donar a aquest document⁵.

El supòsit bàsic que la tècnica del *matching* necessita per obtenir estimacions vàlides de l'impacte és que, més enllà de les variables que s'hagin tingut en compte per realitzar els emparellaments, no hi ha cap altre factor que tingui influència simultàniament sobre el procés de participació en el programa i sobre els *outcomes* d'interès. O, dit amb altres paraules, la hipòtesi

fonamental és que tot el biaix de selecció en la participació al programa es produeix en variables sobre les quals tenim informació, de tal manera que en tenir-les en compte a l'hora de construir el grup de comparació, la simple diferència d'*outcomes* entre participants i no participants "emparellats" constitueix una mesura vàlida de l'impacte del programa. Lògicament, quan més ampli sigui el conjunt de variables sobre les que es disposa d'informació, més plausible resultarà defensar que hem aconseguit reduir el problema del biaix de selecció.

De fet, fruit de la riquesa i exhaustivitat de les bases de dades disponibles en l'àmbit de les PAO, no és estrany que el matching hagi estat profusament utilitzat arreu per avaluar-ne l'impacte en els darrers anys (Kluve, 2010). Tanmateix, no són pocs els analistes que consideren que la identificació basada únicament en variables observables resulta massa feble de cara a eliminar el biaix de selecció, ja que els processos de captació de participants de moltes PAO es caracteritzen, precisament, per un cribratge substancial basat en elements actitudinals o motivacionals, completament absents de les bases de dades. Tot i això, com apunta Vera (2012), alguns investigadors assenyalen que utilitzar informació prèvia sobre l'outcome d'interès (p. ex. la participació laboral), per a diversos períodes anteriors a l'inici del programa, pot ajudar molt a obtenir resultats fiables. En qualsevol cas, ben igual que amb la regressió discontinua, la plausibilitat dels supòsits del matching hauran de ser valorats a la llum de cada aplicació concreta.

3.4 MODEL DE DOBLES DIFERÈNCIES

El model de dobles diferències (DD) pot emprar-se en aquelles situacions en les que un grup d'agents (persones, OTG, municipis...) són exposades a un determinat programa ("tractament") i la resta no. El cas més senzill és el de dos grups i dos períodes: en el primer període, cap dels dos grups participa en el programa; en el segon període, en canvi, un grup hi participa i l'altre no. Un exemple real d'una situació d'aquest estil, avaluada de fet mitjançant un disseny DD, és la que va acompanyar la implementació del New Deal for Young People al Regne Unit, ja que el programa únicament va estar operatiu durant la fase pilot en un conjunt limitat de territoris (Blundell et al., 2004).

L'estratègia d'identificació de l'impacte que proposa un DD consisteix a comparar, d'una banda, la diferència en els *outcomes* entre ambdós grups en algun moment després de la finalització del programa i, de l'altra, la diferència existent abans del seu inici. Fet això, si l'inevitable biaix de selecció entre ambdós grups roman constant al llarg del temps, el canvi entre abans i després en la diferència d'*outcomes* entre el grup de participants i el de comparació (això és, la diferència de les diferències) constitueix una estimació vàlida de l'impacte del programa.

El supòsit clau en un disseny DD és que qualsevol diferència pre-programa entre el grup de tractament i el de comparació s'hagués mantingut constant encara que el programa no hagués existit. Sota aquest supòsit, els *outcomes* post-programa del grup de comparació, degudament

ajustats per el diferencial pre-programa, esdevenen un contrafactual creïble dels outcomes post-programa del grup de tractament. És important adonar-se que, comparat amb el mètode del matching, el supòsit d'un model de DD és menys restrictiu en la mesura en que ambdós grups poden ser diferents en factors inobservables (per exemple, la motivació en el cas d'un programa de formació), sempre i quan aquestes diferències romanguin constants al llarg del temps; en canvi, en el cas del matching, per a que la tècnica proporcioni una mesura vàlida de l'impacte, cal que ambdós grups no difereixin en cap factor no observable. No són pocs els investigadors, com ara Card et al. (2011), que veuen en aquest punt un clar avantatge del model de DD sobre el matching, fins l'extrem que recomanen amb entusiasme la utilització del primer i amb molta més cautela la del segon.

Tanmateix, cal no oblidar que també el model de DD, com la resta de dissenys no experimentals, està subjecte a limitacions si no es satisfan els supòsits que el sustenten. En aquest sentit, el punt feble dels models de DD és la possible existència de factors inobservables que varïïn al llarg del temps. Així, seguint amb l'exemple del programa Segona Oportunitat, si la motivació de tractaments i controls varia al llarg del temps, i no podem observar aquesta variable, no podem estar plenament segurs que aquest factor no sigui la causa de l'evolució diferencial de l'outcome en el grup de tractament respecte del de control i, per tant, que la magnitud de l'impacte estimat per a la política no en sobreestimi el seu efecte real. Així doncs, si volem que els resultats d'una avaluació d'impacte que utilitzi un disseny DD resultin creïbles, haurem d'oferir arguments que permetin descartar l'existència de característiques inobservables que varïen en el temps de forma diferent entre tractaments i controls.

3.5 ELECCIÓ ENTRE MÈTODES

Els apartats anteriors han posat de manifest l'existència de diversos mètodes susceptibles d'ésser utilitzats a l'hora de mirar d'establir l'impacte d'una determinada política. En general, una visió força compartida entre els avaluadors és que no existeix el mètode ideal, és a dir, un tipus de disseny en particular que, independentment de les circumstàncies, s'hauria d'aplicar de forma universal en totes les avaluacions d'impacte. A la pràctica, per tant, els avaluadors es veuen obligats a decidir entre diverses alternatives. Un element obvi que condiona aquestes eleccions és la disponibilitat de temps i de recursos, però n'hi ha d'altres: les característiques del programa, la importància dels resultats i l'ús que s'en vulgui fer, la disponibilitat de dades, etc. Les línies que segueixen tracten breument aquests aspectes, i argumenten a favor de la necessitat d'aproximar-se a l'elecció del mètode amb una mentalitat oberta, eclèctica i desproveïda d'apriorismes excessius.

Hi ha determinades característiques de les polítiques públiques que augmenten les possibilitats de mesurar amb rigor el seu impacte. Una especialment important és la relativa a la seva novetat i, més concretament, a la seva concepció com a prova pilot. En aquests casos, si es reuneixen una sèrie de condicions, com ara que la demanda potencial sigui superior

a l'oferta o existeixin dubtes sobre l'efectivitat del programa, els experiments socials que utilitzen procediments d'assignació aleatoris poden constituir una forma d'avaluació d'impacte a considerar. En qualsevol cas, malgrat l'assignació no es produeixi de manera aleatòria, un programa pilot que s'implanti només en determinades zones geogràfiques obre les portes a dissenys no experimentals (matching o models DD) que utilitzin les àrees no-pilot per construir grups de comparació.

Un altre avantatge de les polítiques noves, es materialitzin o no mitjançant proves pilot, és que permeten la introducció d'elements d'avaluabilitat mentre es desenvolupa la fase de disseny del programa. Com hem mencionat anteriorment, una avaluació d'impacte és, per definició, una avaluació ex-post, però les millors avaluacions d'impacte són aquelles que es planifiquen ex-ante. La possibilitat més extrema és que el propi desplegament de la política es realitzi pensant en l'avaluació, com és el cas d'un experiment social, però de vegades n'hi ha prou amb planificar una bona recollida de dades abans i després de la intervenció, que afecti a sengles mostres de potencials beneficiaris i no beneficiaris, per incrementar enormement les possibilitats d'obtenir estimacions d'impacte creïbles mitjançant mètodes no experimentals.

Hi ha moltes ocasions, però, en què l'impacte que es desitja avaluar no és el d'una política nova. En aquests casos, donat que resulta impossible influir en "clau avaluadora" sobre el disseny del programa, el repte de l'avaluació consisteix a trobar característiques de la política i fonts d'informació que facin possible l'aplicació de les tècniques quasi-experimentals descrites en aquesta guia.

Així doncs, pel que fa a les característiques del programa, cal buscar-hi elements que possibilitin la construcció de contrafactuals: per exemple, si pels motius que sigui un determinat programa té llistes d'espera, els individus que en formen part poden constituir un grup de control natural respecte el qual estimar l'impacte del programa; així mateix, en la mesura en que existeix variabilitat geogràfica en el grau d'implantació d'una política, les unitats territorials que disposin del programa poden comparar-se amb les que no en disposin (els municipis catalans poden constituir, en el cas d'algunes PAO, una font de variabilitat a explorar en aquest sentit).

Les diverses tècniques en què hem centrat la nostra atenció fins el moment són metodologies d'anàlisi quantitatives. No és estrany la preeminència d'aquest tipus d'enfocament en l'avaluació d'impacte, ja que la qüestió fonamental a resoldre, que no és altra que la construcció d'un contrafactual, és de naturalesa bàsicament quantitativa. No obstant això, existeix una percepció creixent per part dels avaluadors que, per tal de millorar la robustesa de l'avaluació d'impacte, resulta recomanable complementar l'anàlisi utilitzant tècniques qualitatives (p. ex. entrevistes en profunditat o grups de discussió)⁶. El valor afegit que pot aportar llur utilització és permetre a l'equip avaluador millorar el seu coneixement sobre les condicions en què realment opera el programa, les perspectives dels seus beneficiaris, i d'altres elements fonamentals a l'hora d'entendre realment el perquè de l'impacte d'una política o programa (o de la seva

absència).

Notes:

³ Fem abstracció de la mesura de l'outcome en si, que pot ser la inserció laboral o altres coses, per tal de concentrar-nos en els aspectes relacionats pròpiament amb la mesura de l'impacte. En el document "Model d'avaluacions del SOC", es tracten diverses qüestions relatives a la mesura dels outcomes en el cas de les PAO.

⁴ Vegeu Casado (2012) per a una anàlisi més detallada de les possibilitats i reptes que suposa l'avaluació experimental de polítiques públiques. D'altra banda, per a conèixer diverses iniciatives orientades a desenvolupar institucionalment aquest tipus d'avaluacions, podeu consultar la panoràmica de Lázaro (2012) al respecte.

⁵ Els lectors interessats en aprofundir en els aspectes més tècnics del matching poden consultar les referències que s'esmenten a la secció "Per saber-ne més" amb la que es tanca la guia.

⁶ Vegeu Sanz (2011) per a una revisió de les principals tècniques d'anàlisi qualitativa i llur aplicació a l'avaluació de polítiques públiques.

4. COM MILLORAR L'AVALUACIÓ DE LES PAO ARA I AQUÍ?

4.1 DISSENYAR LA POLÍTICA I L'AVALUACIÓ ALhora

El moment en el qual es dissenya l'avaluació d'una política acostuma a tenir conseqüències importants sobre la robustesa dels resultats de l'avaluació. En aquest sentit, malgrat tota avaluació d'impacte mesura retrospectivament els efectes d'un programa, algunes d'elles tenen la característica d'haver estat dissenyades abans que el nou programa s'implementés.

En general, les avaluacions d'impacte dissenyades abans de l'inici del programa, que anomenarem prospectives, permeten aportar evidències més rigoroses sobre els efectes d'un programa per dos motius fonamentals: d'una banda, obren la possibilitat d'influir sobre com el programa es porta a terme i, d'aquesta manera, apostar per fórmules d'implementació que generin grups de tractament i comparació que permetin estimar l'impacte de forma més creïble (per exemple, portant a terme una avaluació experimental amb aleatorització); d'altra banda, ofereixen la possibilitat de definir des de l'inici les peces d'informació que faran falta per portar a terme l'avaluació (mesures adequades dels outcomes, de les característiques dels individus, dels costos d'implementació, etc.) i posar els mitjans per generar-les. En canvi, quan l'avaluació es dissenya quan el programa ja es troba en marxa (o molt sovint finalitzat), l'elecció de l'analista es troba fortament condicionada no tant sols per quines siguin les dades disponibles i llur qualitat, sinó també per les característiques d'implementació del programa que, en el pitjor dels casos, poden fins i tot fer impossible l'aplicació de cap de les tècniques esmentades prèviament en no existir la possibilitat de definir un grup de comparació.

Fins ara, totes les avaluacions d'impacte dels programes del SOC que s'ha dut a terme, incloses les d'Àlbum, han estat dissenyades un cop els programes en qüestió ja havien finalitzat. Així doncs, un dels reptes del SOC és incorporar l'avaluació des de l'inici del procés de gestió d'una nova PAO. Això passa per descriure clarament quins dels mètodes d'estimació contrafactual es faran servir (experiment social, dobles diferències, etc.), els outcomes concrets sobre els quals es mesurarà l'impacte (participació laboral, ingressos, etc.), així com les fonts d'informació que s'utilitzaran no tant sols per mesurar els outcomes sinó la resta de variables esmentades prèviament (característiques dels individus, costos d'implementació, etc.).

De fet, més enllà que les avaluacions prospectives permeten estimar l'impacte dels programes de forma més rigorosa, el simple fet de "pensar en com avaluar l'impacte" d'un nou programa té efectes positius sobre el propi disseny substantiu de la intervenció. En concret, obliga als responsables del programa a precisar nítidament quins són els outcomes d'interès sobre els quals la intervenció pretén tenir algun impacte, així com els mecanismes pels quals s'espera que aquests efectes es produeixin. A més a més, enfrontats a la necessitat de substantivar outcomes, mecanismes i impactes esperats, els encarregats de dissenyar el programa tendeixen en major mesura a revisar el contingut i els impactes de les polítiques dutes a terme

a d'altres països, la qual cosa resulta positiva en la mesura en que s'aprofita el coneixement acumulat arreu.

L'impuls a la realització d'avaluacions prospectives comença a impregnar les activitats d'institucions diverses. És el cas, per posar un exemple, del Banco Iberoamericano de Desarrollo (BID), el qual ha incorporat entre els ítems de prioritització dels projectes susceptibles de rebre finançament la obligatorietat de dissenyar l'avaluació d'impacte de la intervenció. Algunes iniciatives recents de la Comissió Europea, com ara les convocatòries PROGRESS d'impuls a la realització d'avaluacions d'impacte de les PAO, denoten un clar interès en l'entorn europeu per avançar en aquesta mateixa direcció.

4.2 LES EXPERIÈNCIES PILOT: NO MÉS OCASIONS PERDUDES

Les anomenades experiències pilot representen un àmbit en el qual la realització d'avaluacions prospectives esdevé, a priori, quelcom especialment senzill d'impulsar. En el cas de les PAO, però també en el d'altres intervencions educatives o dels serveis socials, no és estrany a casa nostra proposar una implementació a petita escala d'un nou programa o política, amb el propòsit precisament de comprovar si es tracta o no d'una bona idea. Tanmateix, tret d'alguna excepció puntual, la majoria d'aquestes experiències pilot no han definit una estratègia avaluativa que permeti estimar amb un mínim de rigor l'impacte de la intervenció en qüestió (això és, proposant l'ús d'alguna de les tècniques descrites a l'apartat 3 d'aquesta guia). En general, sembla que el principal interès d'aquestes experiències pilot sigui testar elements relatius a la implementació del programa, incloent llur viabilitat "política", per tal de millorar-ne el seu disseny abans de procedir a la seva generalització, la qual sovint ja ha estat malauradament decidida abans d'iniciar el pilot.

Tanmateix, la majoria de pilots presenten, per naturalesa, algunes característiques molt propícies per a la realització d'avaluacions prospectives. En primer lloc, atès que per definició un pilot només s'aplica a una fracció dels potencials beneficiaris -ja siguin persones, entitats o municipis-, l'existència d'un grup de comparació està sempre garantida. De fet, donat que tothom accepta que si es porta a terme un pilot hi haurà potencials beneficiaris que no participaran en el programa, s'obre la porta a emprar l'aleatorització com a mecanisme d'assignació, ja que l'atzar acostuma a percebre's en aquestes situacions com un criteri just de selecció. En definitiva, si per exemple el SOC tirés endavant un pilot per testar l'efectivitat d'un nou dispositiu d'assessorament per als aturats, i hi hagués pressupost per finançar llur implementació només en 15 OTG, podria optar-se per seleccionar aleatòriament les OTG participants i així avaluar experimentalment l'impacte de la intervenció.

D'altra banda, encara que no s'aleatoritzi quins dels potencials beneficiaris es converteixen en participants, la lògica subjacent del pilot es que s'està "provant" quelcom i, per tant, això facilita la introducció de consideracions avaluatives en el disseny del programa. En particular

i continuant amb l'exemple anterior, si la previsió fos introduir el nou dispositiu en 15 OTG no triades a l'atzar, encara seria possible dissenyar per avançat una avaluació d'impacte basada en un model de dobles diferències que comparés abans i després l'evolució dels outcomes d'interès a les 15 OTG participants i a les 56 que no hi participen. El disseny de l'avaluació inclouria una previsió sobre tota la informació que caldria recopilar per tal de maximitzar les possibilitats d'obtenir una estimació robusta de l'impacte del programa.

4.3 L'EXCÉS DE DEMANDA I L'ALEATORITZACIÓ COM A MECANISME D'ASSIGNACIÓ JUST

És poc habitual que els responsables d'una determinada PAO, ja siguin els plans d'ocupació, la formació ocupacional o el programa Suma't, reconeixin obertament l'existència d'un excés de demanda. Es tracta d'un posicionament comprensible, atès que l'excés de demanda s'interpreta automàticament per part de la opinió pública com una senyal d'insuficiència de recursos. Tanmateix, malgrat és cert que l'excés de demanda només ens hauria de preocupar en la mesura que contingui persones que necessiten el programa, costa de creure que amb taxes de cobertura com les que s'observen en alguns programes, per sota del 10% dels aturats elegibles (Suma't o NCNO), no hi hagi aturats que necessiten el programa però que no hi accedeixen perquè les places es troben restringides.

La reacció per part de les entitats proveïdores de PAO amb excés de demanda és portar a terme processos estrictes de selecció entre els candidats elegibles: així doncs, d'entre tots aquells que compleixen els requisits formals de la convocatòria, només hi acaben participant els que a més a més demostren estar més motivats, semblen tenir més aptituds, etc. En aquest sentit, sense entrar a discutir si realment aquests criteris de selecció són els més adequats o no, el que proposem és fer-los explícits, mitjançant sistemes de baremació que de fet ja utilitzen informalment moltes de les entitats, i fer sorteigs allà on hi hagi diversos candidats que es troben empatats a punts.

Una alternativa més radical, en la qual les entitats proveïdores perdrien bona part de la seva discrecionalitat actual, seria fer un sorteig entre tots aquells aturats que, d'una banda, manifestessin interès a participar en el programa i, de l'altra, reunissin els requisits previstos a la convocatòria. L'assignació aleatòria de la participació podria considerar-se un criteri de priorització just si, com fora desitjable, els criteris de la convocatòria capturessin adequadament la necessitat (o intensitat de la problemàtica) dels aturats als quals el programa pretén ajudar.

4.4 ELS SUPLENTS: FENT DE LA NECESSITAT VIRTUT

L'existència de suplents és habitual en la majoria de programes del SOC que tenen un component formatiu. Aquest és el cas lògicament de la formació ocupacional (FOAP), però també d'altres programes on es contemplen accions formatives de naturalesa diversa, com és el cas de Suma't o de Noves Cases per Nous Oficis, per posar només dos exemples. Els suplents es

caracteritzen per reunir els requisits de la convocatòria, i per haver superat el procés de selecció addicional al qual ens hem referit abans, però finalment no han iniciat la participació en el programa perquè hi havia d'altres candidats més adequats o que havien arribat una mica abans. Tanmateix, si finalment es produeixen baixes entre els titulars, alguns dels suplents són novament contactats i es converteixen en participants.

El fet que hi hagi suplents s'explica per la concurrència simultània de dos factors. En primer lloc, són el resultat de l'existència d'un excés de demanda que afecta a persones que el necessiten, els interessa i se'n podrien beneficiar, ja que els propis gestors del programa n'accepten la seva participació ex-post si es produeixen vacants. En segon lloc, precisament perquè la seva raó de ser és la cobertura de vacants, constitueixen un mecanisme d'assegurament per als gestors del programa contra dos tipus de contingència: d'una banda, "errors" en els mecanismes de selecció de participants, en el sentit que alguns dels participants titulars abandonen el programa per una manca d'interès o motivació que no va ser correctament detectada; i, d'altra banda, abandonaments per condicions sobrevingudes, no anticipables pels gestors, com ara el fet de trobar una feina, un trasllat de la família, etc. Les entitats tenen interès a disposar de suplents com a "mecanisme d'assegurament" perquè, si no cobrissin les vacants que es generen, se'n ressentiria l'indicador de gestió que valora com a negatiu un percentatge elevat d'abandonaments.

En qualsevol cas, més enllà del perquè de la seva existència, la qüestió important des de la perspectiva de l'avaluació és que el fet que hi hagi suplents obre la porta a estratègies d'estimació de l'impacte molt interessants. En particular, si restringim ara el nostre focus d'atenció exclusivament en els suplents, el mecanisme de selecció de participants entre els suplents "s'assembla" al d'un sorteig: si el suplent està inscrit en un curs on es produeixen moltes (poques) vacants, la seva probabilitat de participar és més alta (baixa); i, atès que no està en mans dels individus alterar el número de "tiquets" que tenen per a la rifa, les característiques d'aquells suplents que acaben participant són idèntiques d'aquells que finalment no ho fan (incloent els atributs no observables).

La nostra proposta és augmentar encara més l'atractiu avaluatiu de l'existència de suplents mitjançant un protocol explícit del SOC que, a grans trets, proposés als proveïdors seguir els passos següents:

- 1.** Definir un percentatge obligatori de suplents (p.ex. 1 per cada 3 places) per a cada convocatòria. En el cas que l'entitat no assolís aquest percentatge, hauria de justificar el perquè no s'hi ha arribat.
- 2.** La definició de "suplent" hauria de ser dicotòmica (sí o no), sense cap sistema de punts que ordenés entre si els suplents. Es podrien utilitzar barems per classificar als candidats, però un cop definit el llindar que marca l'inici de la condició de suplent, la puntuació de l'individu esdevé irrellevant.

3. La tria dels suplents per convertir-se en participants hauria de basar-se en un sorteig.

Tanmateix, si s'opta per no voler restringir la capacitat de les entitats per "ordenar" els suplents i basar el criteri de repesca en aquestes ordenacions, caldria exigir llavors que el barem i les puntuacions fossin conegudes pel SOC. Això permetria intentar estimar l'impacte mitjançant la tècnica de la regressió discontinua comentada a l'apartat 3, tot i que les limitacions en aquest cas podrien ser importants.

4.5 LA MILLORA DE LES BASES DE DADES

L'abast i qualitat de les dades disponibles són un dels principals condicionants a l'hora de poder portar a terme una bona avaluació d'impacte. En el cas de les PAO, tal i com palesen les diverses avaluacions realitzades per Ivàlua (2009, 2011 i 2013), el punt de partida és força millor que el que s'observa en d'altres àmbits d'actuació pública a Catalunya. En particular, existeixen tres bases de dades que poden emprar-se per estimar quantitativament l'impacte laboral de les PAO finançades pel SOC⁷:

1. **Base de dades d'integració del SOC (SIPAO).** Identifica els participants de totes les PAO finançades pel SOC, així com cadascuna de les entitats proveïdores de programes (ens locals, empreses, entitat sense afany de lucre, etc.). Inclou la data d'inici i finalització de la participació en el programa per part dels beneficiaris, així com d'altres variables que permeten "qualificar" la naturalesa de la participació (abandonament, motiu de l'abandonament, grau d'aprofitament, etc.). Així mateix, quan el programa inclou projectes de naturalesa diversa, el SIPAO pot aportar informació, per exemple, sobre el sector d'activitat de l'acció formativa o sobre el tipus d'empresa en el qual es desenvolupen les pràctiques.
2. **Base de dades SICAS-SISPE.** És la base de dades general dels demandants d'ocupació de Catalunya. Els registres històrics corresponen al darrer dia laborable de cada mes, des del maig de 2005. Conté informació socio-demogràfica bàsica (sexe, any de naixement, edat, nacionalitat, població de residència i nivell formatiu), sobre les preferències i disposició a treballar (restriccions de jornada, disposició a treballar fora del municipi o comarca, número d'ocupacions diferents demanades i tipus d'ocupació demanada), història d'atur (data d'inici del període d'atur actual), història laboral (declaració del temps total treball i declaració del sector de la darrera ocupació) i altres atributs rellevants per a l'ocupació (percepció de la prestació activa, coneixement d'idiomes, declaració de discapacitats, etc.). No conté informació sobre algunes característiques rellevants, com ara l'estructura de la llar (fills menors, edat del fill més petit, monoparentalitat, etc.), la història completa d'atur (número i durada dels períodes anteriors d'atur), història laboral anterior a 2005 o el temps que manca per l'exhauriment de la prestació o subsidi d'atur.
3. **Base de dades de contractes laborals.** Estructurada en arxius mensuals i conté els contractes

i pròrroques realitzats en el mes de referència. Dóna informació tant de les característiques sociodemogràfiques bàsiques de la persona contractada com de l'empresa contractant i del tipus de contracte.

4. Base de dades d'afiliació a la Seguretat Social. Estructurada en arxius trimestrals, corresponents al darrer dia laborable de cada trimestre, conté informació sobre les altes a la Seguretat Social, incloent el tipus de contracte (la qual cosa permet identificar si és indefinit o temporal) i el DNI de l'entitat contractant (que permet discriminar si és privada, pública o entitat sense ànim de lucre). Com succeeix amb les altres bases de dades, els registres històrics estan disponibles des de 2005.

La feina que cal portar a terme amb aquestes quatre bases de dades per tal d'avaluar l'impacte d'una determinada PAO, ja sigui el programa Suma't o els Plans d'Ocupació, és sempre la mateixa i consisteix a construir una sola base de dades que contingui: i) tots els aturats de Catalunya que estaven registrats a les OTG en el moment d'iniciar-se el reclutament de participants (SICAS); ii) una variable que distingeix quins d'aquests aturats varen participar en el programa objecte de l'avaluació, quins en d'altres PAO i quins en cap (SIPAO); iii) diverses variables relacionades amb la participació en el programa, com ara la durada, l'abandonament, el tipus d'activitats, etc. (SIPAO); iv) les característiques socio-demogràfiques de totes les persones del punt i (SICAS); i finalment, v) els outcomes laborals de totes les persones del punt i, hagin o no participat en el programa, des de la seva data de finalització fins el darrer trimestre disponible en els talls trimestrals d'afiliació (Seguretat Social) així com via els fitxers de contractes. El NIF-NIE dels individus és el camp que permet enllaçar les quatre bases de dades entre si (SIPAO, SICAS, SS i contractes) per tal d'obtenir la base de dades analítica que acabem de descriure.

El principal problema del procediment anterior és que ha calgut repetir-lo, ex profeso, per a cadascuna de les avaluacions d'impacte encarregades pel SOC tant a Ivàlua com a d'altres institucions. Així doncs, una primera millora en l'àmbit dels sistemes d'informació consistiria a desenvolupar una base de dades integrada per a l'avaluació de les PAO, a partir de l'enllaç de SICAS, SIPAO, Seguretat Social, i contractes en un primer moment, i de les dades d'Ensenyament en una segona fase. En aquesta línia el SOC i l'Observatori del Departament d'Empresa i Ocupació fa temps hi estan treballant; de fet, actualment ja estan integrats en el Datawarehouse del Departament els fitxers de l'àmbit laboral. Seria molt beneficiós per incrementar les condicions d'avaluabilitat de les PAO accelerar la fase actual del projecte que és la d'integrar la informació de PAO. Tot i així, hi ha un important escull a superar, la gestió de les PAO a Catalunya no ha comptat mai amb un sistema bo i general (que cobreixi totes les PAO) d'informació per a la gestió i això limita la disponibilitat i qualitat de les dades. En aquest sentit, seria també molt important accelerar els treballs al respecte contemplats al Pla director TIC del SOC. A més a més, seguint l'estela d'altres països del nostre entorn, com ara Alemanya (Kruppe et al., 2008), un projecte d'aquestes característiques hauria de contemplar que la comunitat investigadora pogués emprar la base de dades a voluntat, amb les cauteles legals oportunes, ja

que d'aquesta manera es maximitza la generació de coneixement i es preserven les garanties d'independència en l'anàlisi.

D'altra banda, amb independència que s'acabi o no apostant per la generació d'una base de dades integrada com la que acabem d'esbossar, hi hauria d'altres aspectes amb relació a l'abast i qualitat de la informació que també caldria tractar de millorar:

- Incorporar més variables a SICAS que poden resultar d'interès a l'hora de millorar la comparabilitat entre participants i no participants, especialment si s'empra la tècnica del "matching", com per exemple: l'estat civil, el número de fills i llur edats, l'estat de salut, etc.
- Enllaçar amb d'altres bases de dades administratives, per tal de millorar la informació disponible sobre la percepció de prestacions i subsidis d'atur (INEM) o de la quantia dels salaris percebuts (IRPF o bases de cotització).
- Explorar la possibilitat de realitzar enquestes sobre mostres d'aturats per tal d'enriquir la informació administrativa disponible. En el cas d'Alemanya, per exemple, existeix des de l'any 2008 una enquesta realitzada sobre una mostra de 20.000 aturats, extrets de la base de dades integrada a la que ens hem referit prèviament, que aporta informació sobre atributs rellevants a l'hora d'avaluar l'impacte de les PAO (Caliendo et al., 2010): actituds, factors psicològics, abast de les xarxes socials, intensitat de la cerca de feina, etc.

4.6 MÉS ENLLÀ DE L'IMPACTE: L'AVALUACIÓ ECONÒMICA DE LES PAO

L'avaluació d'impacte ens permet esbrinar l'efectivitat de les PAO, és a dir, si aquells que hi participen milloren (o no) la seva situació a causa del programa. Així doncs, si davant d'una determinada problemàtica, com ara la baixa ocupabilitat dels joves sense ESO, hem posat en marxa diverses polítiques que hi fan front mitjançant aproximacions alternatives (Suma't, NCNO, PQPI, etc.), l'avaluació d'impacte ens permet valorar, com a mínim, si cadascun d'aquests programes resulta efectiu pel que fa a la millora de la inserció laboral, l'increment del capital formatiu, etc.

Tanmateix, atès que els recursos disponibles per a PAO han estat, són i seran limitats, cal donar un pas més i esbrinar si el pressupost assignat a cadascun dels programes, siguin o no per a joves, permet maximitzar l'impacte social dels diners que el contribuent ha posat a disposició del SOC per portar a terme PAO. Des d'aquesta perspectiva, que és la pròpia de l'avaluació econòmica, la comparació entre els diversos programes no pot fer-se únicament d'acord a la magnitud de l'impacte aconseguit en cada cas, sinó que cal considerar el cost dels recursos esmerçats. En particular, la major o menor eficiència dels programes depèn per tant de la relació que existeixi entre, d'una banda, els efectes positius sobre els outcomes d'interès (treball, salaris, etc.) i, de l'altra, la quantitat de recursos que ha calgut mobilitzar per aconseguir aquests impactes. Així doncs, l'eficiència és un concepte més ampli que el

d'efectivitat, el qual queda subsumit en aquell: en d'altres paraules, l'efectivitat és una condició necessària però no suficient per a l'eficiència.

L'argumentari anterior fa referència a l'anomenada anàlisi cost-efectivitat (CE), que és un tipus particular d'avaluació econòmica, i ha estat triada per simplificar l'exposició. La senzillesa del mètode rau en què únicament cal expressar en termes monetaris els costos del programa, mentre que els impactes positius poden expressar-se en unitats naturals (insercions laborals, retorns al sistema educatiu, etc.). En canvi, l'avaluació econòmica emprava també d'altres metodologies més exigents, com ara l'Anàlisi Cost-Benefici (ACB), que obligaria a monitoritzar no tant sols els costos de les PAO sinó també els seus impactes positius (anomenats beneficis). El gran avantatge de l'ACB és que no tant sols permet comparar entre si PAO que tenen impactes sobre outcomes diferents, quelcom que el CE no pot fer, sinó que a més a més permet comparar la rendibilitat de les PAO amb d'altres programes d'impuls econòmic o, fins i tot, amb polítiques públiques d'altres àmbits d'intervenció (salut, educació, etc.)⁸.

A data d'avui, la penetració de l'avaluació econòmica ha estat molt minsa en l'àmbit de les PAO a nivell europeu, tal i com destaca Smith (2009) en la seva comparativa entre la UE i els EUA. Tanmateix, cal començar a millorar la informació sobre els costos de les actuacions que es contemplen, iniciant el desenvolupament d'una genuïna comptabilitat analítica de les PAO, ja que el proper pas després de l'exigència d'avaluar l'impacte de les PAO, un cop es consolidi, serà el requisit d'avaluar-ne la seva eficiència. I, si no ho fem per convicció, que fora realment l'assenyat, ho acabarem havent de fer per obligació.

Notes:

⁷ També es possible mesurar l'impacte educatiu d'aquells programes, com ara els PQPI, que explícitament tenen com a objectiu afavorir el retorn dels participants al sistema educatiu. En aquest cas, la informació disponible prové dels registres administratius del Departament d'Ensenyament que permeten identificar, a nivell d'individu (NIF-NIE), diversos outcomes d'interès: matriculació a GES, a Cicles Formatius de Grau Mig, graduació d'ESO, etc.

⁸ Vegeu Raya i Moreno (2013) per a una introducció a l'avaluació econòmica de polítiques públiques.

5. CONCLUSIONS

La situació econòmica i socialment convulsa que travessem constitueix, paradoxalment, una oportunitat propícia per a l'avaluació. La disminució dels recursos disponibles per a la majoria de polítiques públiques, incloses les PAO, fa aflorar espontàniament les preguntes que es troben en l'arrel de l'avaluació d'impacte: quin són els efectes dels diferents programes? N'hi ha de més efectius que d'altres? A més a més, en el cas particular de les PAO, la pressió per profunditzar en el coneixement de l'efectivitat dels programes ha estat més gran, atès que el finançament d'una part de les actuacions en aquest àmbit prové del pressupost comunitari.

Tanmateix, esbrinar quina ha estat l'efectivitat d'una determinada PAO no és una tasca senzilla. En aquest sentit, el principal propòsit de la guia ha estat comunicar a una audiència no especialitzada l'essència de l'avaluació d'impacte: d'una banda, hem establert que mesurar l'impacte d'una intervenció equival a esbrinar si la política ha causat alguna millora en l'outcome d'interès, la qual cosa exigeix determinar quin hagués estat l'outcome dels participants en absència del programa (contrafactual); de l'altra, hem descrit les principals tècniques existents a l'hora d'avaluar l'impacte i com poden emprar-se per avaluar les PAO del SOC. A més a més, hem intentat il·lustrar que hi ha anàlisis, com ara els indicadors de seguiment o les autovaloracions que, tot i la seva popularitat i malgrat proporcionar informació valuosa sobre un programa, no permeten dir res sobre quin ha estat l'impacte d'una determinada política. En síntesi, hem intentat apropar a gestors i polítics, sense sofisticacions tècniques, al cor d'una disciplina d'investigació social aplicada com és l'avaluació d'impacte.

L'avaluació d'impacte, malgrat es basa en l'aplicació de tècniques d'anàlisi pròpies de les ciències socials, no és un exercici acadèmic o, millor dit, no hauria de ser només un exercici acadèmic. La legitimitat de l'avaluació emana de la seva capacitat potencial per millorar l'efectivitat de les polítiques i programes avaluats. Així doncs, no n'hi ha prou amb tenir avaluacions d'impacte tècnicament robustes, realitzades amb bones dades i per investigadors o analistes competents; si es vol que a més a més els resultats de les avaluacions ajudin a millorar el disseny de les PAO, cal que tots els implicats en el seu desenvolupament, ja siguin polítics o tècnics, hagin impulsat decididament la realització de les avaluacions, n'entenguin el contingut i en valorin llur aplicabilitat. La guia que teniu a les mans ha estat escrita, de fet, amb el propòsit d'aconseguir aquesta complicitat activa envers l'avaluació per part dels responsables del SOC.

Som conscients que incrementar el "coneixement avaluatiu" és una condició necessària, però no suficient, per a que els resultats de les avaluacions d'impacte acabin millorant el disseny de les PAO, ja que hi ha múltiples factors (polítics, socials, etc.) que determinen la fesomia d'aquests programes, més enllà de l'efectivitat. De fet, atès que la legitimitat per desenvolupar les PAO resideix en els responsables polítics, el repte no passa en absolut per arraconar la influència d'aquests altres factors i substituir-los pel "veredict" de l'avaluació; ans el contrari,

es tracta més aviat de mesurar adequadament l'efectivitat de les nostres PAO i que siguin els polítics i els gestors els que, legítimament, prenguin les decisions que estimin oportunes, fent cas o no de l'evidència, però disposant d'informació veraç sobre els impactes reals dels programes dels quals en són responsables.

6. PER SABER-NE MÉS

Hi ha diversos documents recents que aspiren, com aquesta guia, a presentar les principals tècniques d'avaluació d'impacte a una audiència no especialitzada:

- Fitzsimons, E.; Vera-Hernández, M. «A practitioner's guide to evaluating the impacts of labor market programs». *World Bank Employment Policy Primer* (December 2009). <http://goo.gl/bzpGg>
- Card, D.; Ibararán, P.; Villa, JM. «Building in an Evaluation Component for Active Labor Market Programs: A Practitioner's Guide ». *IZA DP n° 6085* (October 2011). <http://is.gd/JXuFLO>

D'altra banda, si es busca aprofundir en els detalls tècnics dels diversos mètodes d'avaluació d'impacte, la següent llista ofereix algunes referències d'interès:

(1) Bernal, R.; Peña, X. *Guía práctica para la evaluación de impacto*. Bogotá: Publicaciones CEDE. Universidad de Los Andes, 2011. <http://goo.gl/k6QjT>

(2) Gertler, P. [et al.] *La evaluación de impacto en la práctica*. Washington DC: World Bank Training Series, 2011. <http://goo.gl/uKfs1>

(3) Imbens, G. W.; Wooldridge, J. «Recent Developments in the Econometrics of Program Evaluation». *Journal of Economic Literature* (2009), 47: 5-86.

(4) Burtless, G. «The case for randomized field trials in economic and policy research». *Journal of Economic Perspectives* (1995), 9: 63-84.

Les lectures (1) i (2) són llibres excel·lents sobre el tema de l'avaluació d'impacte. Tots dos utilitzen molts exemples, i estan molt ben documentats, fins i tot contenen les dades i els codis de programes d'ordinador per emprar les tècniques descrites. La lectura (3) és una revisió dels últims avenços metodològics, i la (4) és una discussió molt interessant sobre l'avaluació experimental de polítiques públiques.

BIBLIOGRAFIA

- Blundell, R.; Dias, M. C.; Meghir, C. [et al.] (2004). "Evaluating the employment impact of a mandatory job search program". *Journal of the European Economic Association*, 2(4), 569–606.
- Caliendo, M.; Falk, A.; Kaiser, L. C. [et al.] (2011). "The IZA Evaluation Dataset: towards evidence-based labor policy making". *International Journal of Manpower*, 32(7), 731–752.
- Card, D.; Ibarra, P.; Villa, J. (2011). "Building and evaluation component for active labour market programs: a practitioner's guide". IZA. Retrieved from <http://goo.gl/yR1VM>
- Casado, D. (2012). "Per què no avaluem les polítiques públiques com els fàrmacs? Una aposta per l'experimentació social". *Avaluació per al Bon Govern*, (3). www.avaluació.cat
- Graversen, B. K.; Van Ours, J. C. (2008). "Activating unemployed workers works; Experimental evidence from Denmark". *Economics Letters*, 100(2), 308–310.
- Haynes, L. et al. (2012). Test, Learn and Adapt. Developing Public Policy with Randomised Controlled Trials. Cabinet Office. Behavioural Insights Team. Retrieved from <http://is.gd/U29XII>
- Heckman, J. J.; Smith, J. A. (1998). "Evaluating the welfare state" (No. 6542). Cambridge: *National Bureau of Economic Research*. Retrieved from <http://www.nber.org/papers/w6542>
- Hendra, R.; Riccio, J. A.; Dorsett, R. [et al.] (2011). Breaking the low-pay, no-pay cycle: Final evidence from the UK Employment Retention and Advancement (ERA) demonstration (Vol. 765). Department for Work and Pensions.
- Kluve, J. (2010). "The effectiveness of European active labor market programs". *Labour Economics*, 17(6), 904–918.
- Kruppe, T.; Müller, E.; Wichert, L. [et al.] (2008). "On the definition of employment and its implementation in register data: the case of Germany". *Schmollers Jahrbuch*, 128(3), 461–488.
- Lázaro, B. (2012). "Les noves formes d'inversió social com a motor d'innovació de les polítiques públiques". *Avaluació per al Bon Govern*, (4). www.avaluació.cat
- Raya, J. M.; Moreno, I. (2013). Guia d'introducció a l'avaluació econòmica (En premsa). Barcelona: Ivàlua.

Sanz, J. (2011). *La Metodologia qualitativa en l'avaluació de polítiques públiques* (No. 8). Barcelona: Ivàlua. Retrieved from <http://goo.gl/SXos2>

Schlotter, M.; Schwerdt, G.; & Woessmann, L. (2010). *Methods for causal evaluation of education policies and practices: An econometric toolbox* (No. 4725). Berlin: IZA Institute for the Study of Labour.

Smith, J. (2004). *Evaluating local economic development policies: Theory and practice*. In *Evaluating local economic and employment development: How to assess what works among programmes and policies*. (OCDE., pp. 287–332). Paris.

Smith, J. (2009). *What can the ESF learn from US evaluations of active labour market programs? Presented at the Evaluation and Performance Management of Job Training Programs: What Can the European Social Fund Learn from the WIA Experience?*. Michigan: APPAM.

ANNEX. EXEMPLES D'AVALUACIONS DE POLÍTIQUES ACTIVES D'OCUPACIÓ

EXEMPLE 1. DISSENY EXPERIMENTAL

Impacte de la provisió pública versus privada d'un programa d'orientació i cerca de feina a França

- **Publicació:** Behaghel, L., B. Crépon and M. Gurgand (2012). Private and public provision of counseling to job-seekers: Evidence from a large controlled experiment. IZA Discussion Paper 6518.
- **Mètode:** Disseny experimental
- **Lloc:** 22 regions administratives de França
- **Període:** Gener 2007 - Març 2008
- **Fonts de dades:** Registres administratius acompanyats d'una enquesta telefònica

Context i contingut de la política

Els programes d'orientació i cerca de feina han rebut una atenció especialment important en els darrers anys, sobretot perquè les avaluacions d'aquests programes mostren que són efectius, sobretot si se'ls compara amb d'altres PAO més tradicionals, com ara la formació o l'ocupació subvencionada. A l'inici del programa, la taxa d'atur era només del 8.4% però una part important d'aquests aturats (30%) havien estat a l'atur durant un any o més.

L'any 2006 el fons de prestacions d'atur va contractar a proveïdors privats per a portar a terme una nova intervenció amb la idea d'estalviar diners a la Seguretat Social. En concret, a canvi d'un pagament per aturat, els proveïdors els havien d'orientar i ajudar intensivament en la recerca de feina. Addicionalment, aquest servei privat havia de servir de competència al fins aquell moment monopoli públic i així incrementar la seva eficiència, que estava fortament qüestionada.

Els 11 proveïdors privats (bàsicament agències de treball temporal) van ser escollits mitjançant un procés de subhasta a cadascuna de les 22 regions participants. L'esquema de pagament als proveïdors era igual a totes elles: un 30% a l'inici, un 35% si l'aturat trigava menys de 6 mesos a trobar una feina, i un 35% si la mantenia durant més de 6 mesos. Tanmateix fruit de les diferències en les subhastes, l'interval de pagament per aturat es va situar entre els 3.000 i els 4.000 euros en els casos més favorables, i entre els 900 i els 1.200 euros en els més desfavorables.

La subcontractació de serveis públics a intermediaris privats pot generar problemes, sobretot quan la qualitat o l'abast de les accions no es poden observar adequadament. En el cas dels serveis d'orientació i cerca de feina, l'heterogeneïtat dels treballadors fa que dimensions com l'esforç o les habilitats de cadascun dels aturats siguin molt difícils d'observar. L'estructura del contracte amb els proveïdors privats pot generar, per tant, incentius perversos que contrarestin el potencial d'eficiència de les empreses privades: d'una banda, fent "cream-skimming" amb els aturats, és a dir, només reclutar aquells amb bones possibilitats de trobar feina; i, de l'altra, "aparcant" als aturats més difícils d'inserir, donant-los-hi només un mínim de serveis.

Context i disseny de l'avaluació

Durant el 2006, el fons de Seguretat Social va assegurar que l'experiència amb els agents privats havia estat molt exitosa. Malauradament, aquesta afirmació es va basar en indicadors d'ocupació sense considerar la situació contrafactual, quelcom arriscat atès el possible biaix de selecció derivat del fet que només un 30% dels aturats assignats als proveïdors privats varen acabar participant. Així les coses, amb el propòsit d'obtenir una mesura creïble de l'impacte del programa, es va portar a terme una avaluació experimental amb assignació aleatòria de la participació.

L'avaluació volia contestar dues preguntes. En primer lloc, buscava conèixer si el mecanisme d'orientació intensiva (cada treballador social té un màxim de 40 aturats assignats i com a mínim un contacte setmanal -via e-mail o telèfon- i un mensual -cara a cara- amb l'aturat) era efectiu respecte a l'itinerari anterior (120 aturats per treballador amb un únic contacte mensual). L'altre pregunta que interessava respondre era si el servei privat era efectivament més efectiu que el servei públic.

Donat que els proveïdors privats només gestionaven un subconjunt dels aturats de l'agència pública, concretament aquells que portaven menys de tres mesos a l'atur i que percebiien una prestació d'atur, l'avaluació es va concentrar únicament en aquest grup. Així, quan un aturat d'aquestes característiques assistia per primer cop a la seva corresponent oficina d'ocupació, un cop analitzat el perfil, es carregava a l'usuari en una extranet que assignava aleatòriament a l'aturat a un dels tres grups següents: prestador privat (grup de tractament 1), prestador públic amb itinerari intensiu (grup de tractament 2) o el servei públic tradicional (grup de tractament 3). Així doncs, el disseny de l'avaluació no tant sols permet analitzar l'efecte de la intervenció intensiva (2 vs 3), sinó també comparar l'eficiència dels agents privats amb la del servei públic quan atenen perfils similars d'aturats (1 vs 2).

Resultats

En primer lloc, d'acord amb els resultats de l'avaluació, l'orientació intensiva en la cerca de feina incrementa la transició cap a l'ocupació tant quan es porta a terme en el servei públic com

quan la realitzen proveïdors privats. En concret, la taxa de sortida de l'atur s'incrementa de 4 a 9 punts percentuals respecte a la intervenció tradicional.

En segon lloc, els impactes són consistentment més alts per als participants que van ser atesos per proveïdors públics. L'anàlisi qualitativa explica aquest resultat pel fet que els proveïdors privats "aparquen" aquells aturats de col·locació més difícil, quelcom que els autors relacionen amb l'elevada quantia del pagament inicial (de 900 a 1.200 euros). Una evidència d'això, constatada a partir de l'enquesta telefònica, és que els proveïdors privats tracten de manera diferenciada les persones d'acord al seu perfil. Per exemple, els proveïdors privats tenen més contactes amb alguns participants i menys amb d'altres, mentre que a l'agència pública els contactes i entrevistes són més uniformes.

EXEMPLE 2. REGRESSIÓ DISCONTÍNUA

La millora de la inserció laboral dels joves aturats al Regne Unit

- **Publicació:** De Giorgi, Giacomo (2005). *The New Deal for Young People Five Years On*. "Fiscal Studies", vol 26 (3) 371-383.
- **Mètode:** Regressió discontinua
- **Lloc:** Regne Unit
- **Període:** Abril 1998 -Desembre 2003
- **Fonts de dades:** Registres administratius del programa (New Deal Evaluation Database, NDED) combinats amb el registre d'aturats del Regne Unit (JUVOS)

Context i disseny de la política

Durant la dècada passada al Regne Unit es van començar a implementar una sèrie de polítiques destinades a incrementar l'ocupabilitat i habilitats dels beneficiaris de programes socials, alhora que es lligava la percepció de les prestacions a la participació en polítiques d'activació. Un d'aquests programes és el *New Deal for Young People*, de participació obligatòria, destinat als joves d'entre 18 i 24 anys que han estat a l'atur durant al menys 6 mesos i que reben una prestació d'atur.

Aquest programa consta de dues etapes. Durant la primera etapa, que dura quatre mesos, s'acompanya a l'aturat en la cerca intensiva de feina, formació en habilitats bàsiques i confecció d'un CV. Aquesta etapa també inclou reunions amb el mentor un cop cada dos setmanes en la qual els joves han de reportar quines activitats han estat fent de cara a trobar feina. En cas de que aquesta etapa no sigui exitosa i que l'aturat no trobi feina, s'activa la segona etapa

durant la qual, de comú acord amb el treballador social, es poden triar diferents opcions. Una primera opció passa per col·locar al jove en una feina subvencionada durant sis mesos. Una segona opció, més adient per als joves sense formació bàsica, són els cursos de formació de 12 mesos de durada, durant els quals el jove continuava cobrant la prestació d'atur. La tercera opció era una feina de voluntari on el participant durant sis mesos rebia la prestació d'atur i un complement. Finalment, la quarta opció era una feina en el sector públic, pensada per als perfils més difícils de col·locar. Addicionalment, acabada la segona etapa, el programa es podia estendre durant 3 mesos més amb unes accions molt semblants a les de la primera etapa.

Context i disseny de l'avaluació

El programa es va començar a aplicar l'any 1998 i des d'aquell moment fins a desembre de 2003 hi havien participat més d'un milió de joves, un 75% dels quals eren homes. El pilot d'aquest programa ja s'havia avaluat prèviament, amb uns resultats positius: el programa incrementava la probabilitat de trobar feina en 10 punts percentuals. L'avaluació que ara comentem, en canvi, intenta mirar l'efecte del programa un cop havia estat completament implementat, per tal de comprovar si els bons resultats de la fase pilot es mantienien o no.

En particular, de cara a identificar correctament l'impacte del programa, es va utilitzar una discontinuïtat existent en el seu disseny: únicament els joves de menys de 25 anys hi podien participar. Així, sota el supòsit que els individus que es troben al voltant dels 25 anys d'edat tenen una evolució similar, aquesta discontinuïtat ens permet mesurar (localment) l'efecte de la política, ja que les persones que per exemple tenen 24 anys i 11 mesos estaran obligades a participar, mentre que les que tenen 25 anys i un mes no ho faran. Es pot assumir, no obstant això, que aquest llindar d'edat és bàsicament arbitrari i que, per tant, sembla plausible pensar que no té cap altra conseqüència en el mercat laboral. De fet, quan es comparen altres característiques individuals, com per exemple l'experiència laboral, el nivell educatiu, la nacionalitat, etc., els dos grups són pràcticament idèntics.

Resultats

La intervenció es defineix com el total d'etapes del programa i no com cadascuna de les diferents opcions de la segona etapa. L'outcome d'interès per avaluar l'impacte és la probabilitat de trobar feina durant els 18 mesos posteriors a l'inici del programa (és a dir, un cop esgotats els 6 mesos d'atur).

Els resultats del programa són positius segons l'avaluació. En particular, com a conseqüència de participar en el programa, els joves de menys de 25 anys tenen una probabilitat de trobar feina durant els 18 mesos posteriors 6 punts percentuals més alta que d'altres joves, amb les mateixes característiques però just per sobre dels 25 anys. L'efecte positiu resulta, a més a més, relativament estable en el temps. Així doncs, un cop implementat completament, el

programa segueix tenint els efectes positius que ja havia detectat l'avaluació portada a terme durant la fase pilot.

EXEMPLE 3. MATCHING

Els efectes a Alemanya de les polítiques actives d'ocupació per a joves aturats

- **Publicació:** Caliendo, M., S. Künn and R. Schmidl (2011). *Fighting youth unemployment: The effects of active labor market policies*. IZA Discussion Paper 6222.
- **Mètode:** Matching
- **Lloc:** Alemanya
- **Període:** 2002- 2008
- **Fonts de dades:** Registres administratius de l'IAB

Context i disseny de la política

En entrar al mercat laboral, els joves normalment troben massa obstacles per trobar feina, la qual cosa es veu reflectida en una probabilitat de ser a l'atur significativament més alta que la de la mitjana de la població. En aquest context, degut a l'abast de conseqüències negatives que té que una part considerable de la població estigui a l'atur durant llargs períodes de temps (entre altres, una reducció permanent de la probabilitat de trobar feina, un increment dels delictes, una reducció del capital social, etc.), els governs de la majoria de països europeus han fet una despesa considerable en polítiques actives d'ocupació per a joves, amb la intenció d'integrar-los al mercat laboral i/o afavorir el seu retorn al sistema educatiu.

A Alemanya, des de l'any 2000 al 2010, un total de 1,4 milions de joves a l'any han participat en polítiques d'activació, tant en aquelles destinades a millorar l'oferta laboral (per exemple, assistència en la cerca de feina, formació ocupacional de curta i mitja durada, mesures destinades a promoure la participació en formació ocupacional i programes de creació de treball), com en aquelles altres destinades a incentivar la demanda d'ocupació (per exemple, polítiques de subsidi a la contractació i programes de becaris). L'objectiu d'aquesta avaluació era conèixer quin impacte havien tingut en el curt, mitjà i llarg termini, el conjunt de polítiques actives d'ocupació per a joves, tant pel que fa a la integració al mercat laboral com al retorn a l'educació formal.

Context i disseny de l'avaluació

L'avaluació de l'impacte del programa es va fer per als joves de menys de 25 anys que es van registrar a l'atur durant l'any 2002. Els participants són tots aquells que varen participar

en alguna de les polítiques d'activació; al seu torn, els no participants que serveixen com a potencials controls són aquells aturats que, durant els primers dotze mesos de ser a l'atur, encara no havien participat en cap PAO. Finalment, per decidir quins d'aquests aturats actuaven com a controls va utilitzar-se la metodologia de matching o emparellament. En concret, les variables que van considerar-se en el matching van ser les següents:

- A)** característiques soci-demogràfiques: sexe, edat, estat civil, situació familiar i nacionalitat;
- B)** nivell educatiu i estat de salut;
- C)** informació sobre la darrera activitat laboral: experiència laboral prèvia, tipus de contracte en la darrera activitat laboral abans de ser a l'atur, grup ocupacional de l'última ocupació, ingrés diari en la darrera ocupació;
- D)** historial laboral dels darrers tres anys: mesos a l'atur, mesos treballant, mesos en una PAO, mesos en inactivitat, mesos amb un contracte de temps parcial;
- E)** informació sobre la condició actual d'aturat i informació sobre el treballador social assignat: mesos restants de prestació d'atur, mesos des de l'últim contacte amb l'oficina de treball, informació sobre preferències ocupacionals, nombre d'ofertes proposades pel treballador social;
- F)** característiques regionals: taxa d'atur i el creixement del PIB en el darrer any

El total de participants avaluats va ser de 12.400 persones, els no participants de 38.639, i el període d'avaluació va ser de 60 mesos a partir de l'inici de la participació en cadascun dels programes.

Resultats

En termes de millorar les possibilitats de trobar feina dels joves aturats, la perspectiva del conjunt de polítiques d'activació és positiva, ja que l'avaluació troba evidència d'un efecte persistent i estable en gairebé tots els tipus de mesures considerades. Els programes que tenen un efecte més important a llarg termini sobre la participació en el mercat laboral són els subsidis salarials (10 a 20 pp) seguits de l'assistència en la cerca de feina i la formació de curta i mitja durada (5 a 10 pp). Per contra, els programes de creació d'ocupació pública redueixen la probabilitat de trobar feina a mitjà termini. Finalment, pel que fa al retorn al sistema educatiu, i més concretament als itineraris de formació professional, els programes de becaris semblen tenir un impacte positiu sobretot en a mitjà termini.

En general, tot i ser efectives, l'evidència mostra que les polítiques més efectives estan dissenyades per a persones amb un nivell educatiu significativament superior al de la

mitjana dels aturats. En aquest sentit, cal remarcar que únicament les polítiques d'incentiu a la demanda (subsidis salarials) son efectives per als joves amb un nivell educatiu més baix. Tanmateix, com assenyalen els mateixos autors, aquest col·lectiu és precisament el més nombrós en termes d'aturats i, per tant, el que hauria de centrar en major mesura el disseny de les PAO per a joves.

EXEMPLE 4. MODEL DE DOBLES DIFERÈNCIES

Redueixen els subsidis l'atur entre les persones de més edat?

- **Publicació:** Boockmann, B., T. Zwick and A. Ammermüller (2012) "Do hiring subsidies reduce unemployment among older workers? Evidence from natural experiments". *Journal of the European Economic Association*, vol 10 (4) 735-764.
- **Mètode:** Dobles diferències
- **Lloc:** Alemanya
- **Període:** 2001 - 2004
- **Fonts de dades:** Registres administratius de l'IAB

Context i disseny de la política

Els treballadors de més edat normalment tenen moltes dificultats per trobar feina en cas de quedar-se sense. L'any 1998, amb la intenció de millorar aquest problema, el govern alemany va introduir un subsidi a la contractació de persones de 50 anys o més, sempre i quan portessin com a mínim 6 mesos a l'atur durant els 12 mesos abans de ser contractades. Aquesta subvenció representava l'equivalent a un 50% dels costos laborals homogeneïtzats durant un període de 24 mesos. A partir de l'any 2002, el govern va estendre aquesta subvenció a totes les persones aturades de més de 50 anys.

Context i disseny de l'avaluació

El canvi regulatori permet utilitzar la metodologia de dobles diferències per estimar l'efecte dels subsidis sobre la transició a l'ocupació durant els primers 180 dies d'atur. En concret, la reforma va generar dos grups de característiques molt similars: un va quedar afectat pel canvi regulatori (els de 50 anys o més) i l'altre no (menys de 50 anys). En particular, tenint en compte que les persones que portaven més de 6 mesos a l'atur ja tenien dret a ser contractades amb la subvenció, els dos grups d'interès són els aturats de "50 a 50 més sis mesos" i els de "49 a 49 i sis mesos". Així doncs, assumint que aquests dos grups s'haguessin comportat de la mateixa manera en absència del canvi regulatori, la comparació dels dos grups abans i després permet

saber si la transició des de l'atur es va reduir a partir de la introducció de la subvenció. Donat que el canvi es produeix l'any 2002, el primer experiment és analitzat a partir dels aturats entre l'1 d'abril i el 30 de juny del 2001 i del 2002.

Resultats

D'acord amb l'estudi, la probabilitat de deixar l'atur i tornar a l'ocupació no es veu afectada per la disponibilitat de subsidis, excepte per a les dones de l'Alemanya de l'est. Aquesta manca d'efectivitat és deguda principalment als "pesos morts" del programa: els increments en la contractació subvencionada venen acompanyats d'una reducció de les contractacions no subvencionades.

