

Sono utili gli interventi sociali?

Gian Paolo Barbeta

Università Cattolica del Sacro Cuore Milano

Introduzione

Negli ultimi anni, il termine “impatto” ha invaso la discussione sul Terzo settore: impatto sociale, imprese a impatto, finanza a impatto, indicatori di impatto, valutazione di impatto e così via. Ma che cosa vuol dire “impatto”? Siamo davvero capaci di misurarlo? Perché mai dovremmo farlo? Sono alcune delle domande che sorgono spontanee di fronte all'alluvione da cui siamo stati sommersi.

La mia proposta è di non perdere tempo con l'ambiguo sostantivo “impatto” (e i suoi molti possibili significati) per provare invece a rispondere, nella migliore maniera possibile, alla domanda cruciale che dovrebbe porsi chiunque voglia migliorare il sistema di welfare. *Gli interventi, le azioni e le politiche che gli enti del Terzo settore (e le amministrazioni pubbliche) realizzano nel campo del welfare riescono a generare cambiamenti positivi nella vita delle persone?* Questa domanda, apparentemente semplice, in realtà nasconde un'insidia profonda, che diventa immediatamente visibile se riformuliamo leggermente la questione chiedendoci: i cambiamenti che (eventualmente) osserviamo nella vita dei destinatari di interventi e politiche sociali *dipendono* dagli interventi e dalle politiche stesse, oppure si sarebbero realizzati comunque? Siamo dunque in grado di stabilire un nesso di causa-effetto tra gli interventi e i cambiamenti eventualmente osservati?

Se non siamo in grado di rispondere a questa domanda, il resto della discussione sull'impatto è completamente privo di significato e parlare di indicatori, finanza e quant'altro è solo tempo perso.

Rispondere alla domanda è importante perché il nostro sistema economico-sociale attraversa un fase di profondo cambiamento che, se da una parte ha fatto emergere bisogni nuovi (ad esempio quello di sostenere il tasso di natalità, oppure di garantire condizioni di socialità elevata a persone anziane sole), dall'altra continua a mostrare la presenza di antiche necessità (ad esempio quella di ricollocare i lavoratori che hanno perso il loro impiego) che le politiche tradizionali non sono state in grado di affrontare adeguatamente. Dare risposte efficaci a questi bisogni (nuovi e antichi) è una priorità per il Paese, sia per migliorare le condizioni di vita della popolazione, sia per rendere più incisivo, e forse anche meno costoso, il sistema di welfare, concentrandosi sugli interventi “che funzionano” e destinando meglio le risorse che ora sono dedicate a politiche e azioni inutili e forse addirittura dannose.

Identificare ciò che funziona e ciò che non funziona è essenziale anche per superare un antico (e mai discusso) pregiudizio tipico del Terzo settore: gli interventi sociali non possono che fare bene. Purtroppo, esiste ormai un'evidenza diffusa che così non sia: alcuni interventi sono inutili, mentre altri possono persino essere nocivi¹. Pertanto, è importante poter stimare, con rigore e non solo sulla base della moda del momento o dell'opinione personale del proponente, l'efficacia dei diversi interventi e delle diverse politiche realizzate.

Comprendere appieno l'efficacia degli interventi e delle politiche è dunque la condizione necessaria per salvare il welfare italiano, riformandolo nella direzione dell'efficacia basata sull'evidenza.

Che cos'è un effetto?

Chiedersi se gli interventi e le politiche cambiano la vita delle persone significa interrogarsi sui loro “effetti”. È un esercizio piuttosto complesso che, per evitare di produrre risposte inaffidabili, deve essere svolto con rigore, rispettando alcuni passaggi procedurali.

Innanzitutto, ci si deve chiedere che cosa sia un effetto. Credo che tutti possiamo concordare nel ritenere che l'effetto di una azione (di un intervento o di una politica) possa essere definito come “ciò che accade in presenza dell'azione e non sarebbe invece accaduto in sua assenza”. Può apparire una definizione complessa ma è quella che utilizziamo quando, parlando di effetti, proviamo a stabilire nessi di causalità nella vita quotidiana. Diciamo infatti: “Ha avuto accesso al reddito di cittadinanza e pertanto ha potuto far terminare la scuola superiore ai figli” (sottinteso: se non avesse avuto il reddito, i figli non avrebbero potuto studiare). Oppure: “Hai studiato abbastanza e perciò hai superato l'esame” (sottinteso: se non avessi studiato non l'avresti superato). O ancora: “L'ambulanza è stata veloce e il paziente si è salvato” (sottinteso: se fosse stata più lenta il paziente sarebbe deceduto).

In letteratura, questa idea di effetto tratta dalla vita quotidiana prende la definizione di “controfattuale” (o di *outcome* potenziali) e la sua applicazione pratica si deve al lavoro di Donald Rubin (Holland, 1986). Il termine controfattuale deriva dal fatto che la definizione ci porta a ragionare non solo su ciò che accade sotto i nostri occhi (il mondo fattuale), ma anche su che cosa sarebbe accaduto in uno “stato del mondo”

¹ Per un'analisi dettagliata dell'efficacia di interventi in campo educativo, che evidenzia anche l'esistenza di interventi dannosi, si veda per esempio il sito della Education Endowment

Foundation inglese, in particolare la sezione dedicata alle “evidence reviews”: <https://educationendowmentfoundation.org.uk/evidence-summaries/evidence-reviews/>

diverso da quello che stiamo osservando, il cosiddetto mondo “controfattuale” in cui la potenziale causa non è presente: non hai preso il reddito di cittadinanza, oppure non hai studiato o, ancora, l'ambulanza ha viaggiato lentamente.

Effetto su che cosa?

La nozione controfattuale è molto utile perché ci consente di capire come si possa misurare l'effetto di un intervento o di una politica. Tuttavia, per farlo bisogna innanzitutto avere bene in mente quale sia il fenomeno su cui l'intervento o la politica vogliono incidere, cioè il cambiamento che vogliono generare nel mondo. Detto altrimenti, è necessario definire quella che gergalmente viene denominata “variabile di *outcome*”. Negli esempi che stiamo utilizzando, ci interessa capire se alcune azioni (ricevere il reddito di cittadinanza, studiare, disporre di ambulanze veloci) siano in grado di modificare i fenomeni su cui concentriamo la nostra attenzione: la prosecuzione dello studio da parte dei figli, la promozione dello studente, la sopravvivenza del paziente: queste ultime sono dunque le nostre variabili di *outcome*.

Alcune variabili di *outcome* possono essere misurate facilmente e frequentemente, come ad esempio la promozione scolastica; pertanto, se vorremo stimare l'effetto di un doposcuola sulla promozione di coloro che vi hanno partecipato, avremo rapidamente a disposizione un indicatore di successo. In altre circostanze, che dipendono strettamente dagli obiettivi di cambiamento perseguiti dall'intervento il cui effetto vogliamo stimare, le variabili di *outcome* si potranno misurare solo più in là nel tempo; se vogliamo stimare l'effetto della frequenza all'asilo nido sullo stato di salute in età adulta non potremo fare altro che attendere che i bambini crescano².

Se non abbiamo chiaro il cambiamento che vogliamo generare attraverso l'intervento, o se lo possiamo esprimere solo in termini vaghi (ad esempio vogliamo “migliorare il mondo”), non saremo in grado di misurare i valori della variabile di *outcome* e, di conseguenza, neppure l'effetto dell'intervento. In questo caso possiamo parlare di un intervento o una politica mal disegnati. Questo non significa, ovviamente, che l'intervento o la politica non avranno effetto, ma solo che non saremo in grado di produrre una stima precisa dello stesso. In questo caso, ognuno esprimerà le proprie sensazioni, che potranno anche essere molto diverse tra loro perché ognuno interpreterà a modo suo il “cambiamento del mondo”, e non avremo una misura chiara per dire quale sensazione sia più corretta. La logica delle diverse argomentazioni, così come la retorica con cui sono espresse, svolgeranno un ruolo chiave nel convincere e far convergere le opinioni.

A volte invece, pur avendo in mente in modo chiaro il cambiamento che vogliamo ottenere, non avremo modo di misurarlo. Non c'è da spaventarsi di tutto ciò. Molto spesso, nella nostra vita, prendiamo decisioni in condizioni di informazioni insufficienti rispetto agli effetti di qualche cosa. Ad esempio, nello scegliere un chirurgo da cui farsi operare, raramente possiamo contare su indicatori precisi relativi alle

sue prestazioni passate. In questi contesti ci valiamo del giudizio (spesso qualitativo) di qualcuno più competente di noi di cui ci fidiamo, della reputazione del medico nella comunità dei suoi pari, delle sensazioni derivanti da un colloquio, e così via. Non possiamo fare altro e perciò ci accontentiamo.

Ma se possiamo accontentarci nella vita individuale, non dovremmo farlo per gli interventi e le politiche. Infatti, logica e retorica potrebbero non essere una base sufficiente per prendere decisioni rilevanti: proseguiamo l'intervento? Estendiamo la politica?

Misurare l'effetto

Una volta chiarito quale sia il cambiamento che vogliamo generare, per misurare l'effetto di una azione è sufficiente calcolare una semplice differenza (fare una sottrazione) tra la misura della variabile di *outcome* nel caso in cui la causa sia presente (ho avuto il reddito di cittadinanza, ho studiato, l'ambulanza è veloce) e la misura della stessa variabile nel caso in cui la causa sia invece assente (non ho avuto il reddito, non ho studiato, l'ambulanza è lenta). Se percependo il reddito di cittadinanza i miei figli frequentano la scuola, mentre senza reddito di cittadinanza non lo fanno, posso dire che “il reddito di cittadinanza ha l'effetto di far proseguire la carriera scolastica dei miei figli”. Se studiando vengo promosso mentre non studiando vengo bocciato, posso dire che “studiare ha l'effetto di farmi promuovere”. Se quando l'ambulanza arriva velocemente la mia vita è salva, mentre in caso contrario non sopravvivo, posso dire che “un'ambulanza veloce ha l'effetto di salvarmi la vita”.

Tuttavia, per stimare gli effetti delle politiche dobbiamo affrontare un problema empirico rilevante: solo uno dei due “stati del mondo” può essere effettivamente osservato, poiché nessuno può vivere due vite parallele e nessuno può, contemporaneamente, ricevere e non ricevere il reddito di cittadinanza, così come non si può studiare e non studiare, o disporre di un'ambulanza veloce e lenta. Pertanto, la variabile di *outcome* può essere misurata solo in una delle due situazioni, quella fattuale, mentre il suo valore nella situazione controfattuale (non essendo osservabile direttamente) dovrà essere stimato. Se ho avuto il reddito di cittadinanza (situazione fattuale), potrò vedere se i miei figli proseguono negli studi oppure no, ma non potrò mai vedere che cosa sarebbe accaduto loro nel caso in cui non avessi avuto il reddito, e viceversa.

Pertanto, per calcolare l'effetto di un intervento, dovremo trovare un modo affidabile per stimare il valore della variabile di *outcome* nella situazione controfattuale.

Errori nella stima degli effetti

Spesso, per calcolare effetti, si pensa che sia sufficiente disporre di misure delle variabili di *outcome* (indicatori) e ricorrere a semplici calcoli: 1) la differenza *prima-dopo* e 2) la differenza *con-senza*. In entrambe le situazioni, come ve-

² Non si tratta di un esempio teorico. Gli effetti di un intervento realizzato nel Michigan all'inizio degli anni '60 che favoriva la frequenza alla

scuola materna da parte di bambini svantaggiati (il Perry Preschool Program) sullo stato di salute una volta diventati adulti, sono analizzati in un

articolo pubblicato nel 2016, quasi cinquant'anni più tardi, quando i bambini erano effettivamente cresciuti (Conti et al., 2016).

dremo, stiamo facendo ipotesi sulla misura della variabile di *outcome* nella situazione controfattuale e in entrambe le situazioni le ipotesi che implicitamente facciamo rischiano di farci produrre stime distorte dell'effetto. Pertanto, è meglio evitare queste tecniche.

Nel caso della differenza *prima-dopo*, l'effetto viene stimato misurando la variabile di *outcome* prima e dopo la realizzazione dell'intervento e sottraendo il primo valore al secondo. Ad esempio, se il voto in matematica prima del doposcuola era 4 e dopo il doposcuola diventa 6, l'effetto del doposcuola sarebbe quello di aver migliorato il risultato scolastico di due punti (6-4=2). In questa circostanza, stiamo implicitamente supponendo che il voto ottenuto prima del doposcuola rappresenti una buona stima (non distorta) della situazione controfattuale, cioè del voto che si sarebbe ottenuto dopo il doposcuola, nel caso in cui non lo si fosse frequentato. Tuttavia, il voto prima del doposcuola è una stima distorta della situazione controfattuale perché, tra il prima e il dopo, molte cose potrebbero essere cambiate anche senza il doposcuola, e alcune di queste potrebbero avere influito sul risultato in matematica: i genitori potrebbero avere aiutato il figlio, quest'ultimo potrebbe essersi appassionato alla materia, ecc. La situazione prima dell'intervento rappresenta una buona stima di ciò che sarebbe successo dopo l'intervento, nel caso in cui il soggetto non avesse preso parte allo stesso, solo se supponiamo che – nel frattempo – nulla sia cambiato (a parte il doposcuola). Gergalmente diciamo: la situazione prima dell'intervento è una buona stima del controfattuale solo se vale l'ipotesi di "assenza di dinamica spontanea"... ed è generalmente impossibile dire se l'ipotesi valga o meno.

Nel secondo caso, la differenza *con-senza*, l'effetto viene stimato misurando la variabile di *outcome* dopo che l'intervento è stato realizzato, per due gruppi di soggetti (coloro che hanno partecipato all'intervento e coloro che non vi hanno preso parte) e sottraendo la seconda misura alla prima. Ad esempio, se il 75% dei ragazzi che hanno partecipato al doposcuola viene promosso e lo stesso risultato è ottenuto dal 40% di quelli che non vi hanno preso parte, l'effetto del doposcuola sarebbe quello di fare aumentare di 35 punti percentuali (75% - 40%) la probabilità di promozione. Tuttavia, l'esito scolastico dei ragazzi che non hanno partecipato al doposcuola non rappresenta una buona stima della situazione controfattuale, cioè degli esiti che sarebbero stati ottenuti dai frequentanti nel caso di mancata frequenza. Infatti, frequentanti e non frequentanti potrebbero essere sistematicamente diversi tra loro (ad esempio, i primi più volenterosi, i secondi più lazzaroni), sicché l'esito finale potrebbe essere spiegato dalle differenze esistenti prima dell'intervento, piuttosto che dalla frequenza al doposcuola.

La situazione di coloro che non hanno partecipato all'intervento rappresenta una stima non distorta della situazione in cui si sarebbero trovati i partecipanti se non avessero partecipato solo in assenza di quella che gergalmente viene chiamata distorsione da selezione (*selection bias*). Purtroppo, sappiamo bene che uno dei problemi classici delle politiche sociali consiste proprio nella difficoltà di raggiungere i soggetti che, potenzialmente, potrebbero trarre maggiore vantaggio dagli interventi. E sappiamo anche che talvolta rischiamo di erogare i servizi ai soggetti che ne hanno meno bisogno, perché si fanno raggiungere più facilmente rispetto

a quelli in maggiore difficoltà. Per questa ragione, la distorsione da selezione è sempre in agguato.

— Che cosa non si può stimare

Se questi metodi semplici per misurare l'effetto di un'azione o di una politica non funzionano, anzi rischiano di fuorviare la nostra analisi e le nostre stime, che altro possiamo fare?

Dobbiamo dire subito che, se l'intervento o la politica interessano un solo soggetto (o un numero ridotto di soggetti) non c'è modo di ottenere una stima non distorta della situazione controfattuale. In altre parole, non saremo mai in grado di dire che cosa sarebbe successo a Pierino, che ha frequentato il doposcuola ed è stato promosso, se non lo avesse frequentato (o viceversa), perché la vita di Pierino non può ricominciare da capo. Ognuno di noi, conoscendo più o meno bene Pierino e la sua situazione, potrà esprimere il proprio parere, provando a convincere gli altri sulla base della logicità delle sue argomentazioni e della sua capacità di persuasione. Molti lo fanno al bar, il lunedì mattina, spiegando agli amici che se avesse giocato Recoba, l'Inter avrebbe vinto. Ma altri dissenteranno, esprimendo opinioni alternative, e la discussione proseguirà a lungo, spesso senza giungere a una conclusione condivisa.

È interessante perciò osservare che, analogamente al caso di Pierino (e di Recoba), non saremo mai in grado di produrre una stima rigorosa dell'effetto di una organizzazione (ad esempio una cooperativa sociale, misurando il suo "impatto sociale"), nemmeno ammettendo di saper spiegare che cosa sia "l'effetto di una organizzazione". Infatti, non sarà possibile stimare in maniera convincente che cosa sarebbe accaduto al mondo nel caso in cui l'organizzazione non fosse esistita (la situazione controfattuale). Anche in questo caso ognuno argomenterà in maniera più o meno convincente la propria tesi e, probabilmente, continueranno ad esistere pareri opposti. Alcune argomentazioni potranno essere più condivisibili di altre e quindi, ad esempio, molti concorderanno che un'organizzazione che non lega in cantina i tossicodipendenti di cui si occupa sia "migliore" di una che invece lo fa. Tuttavia, così dicendo, non stiamo stimando effetti o impatti, ma semplicemente esprimendoci sulla preferibilità di una pratica all'altra. L'effetto resta insondato.

Poco male. Interrogarsi su azioni che riguardano un singolo caso è di scarso interesse per l'organizzazione complessiva del sistema di welfare, così come è privo di rilievo misurare l'impatto di un'organizzazione.

— Che cosa si può invece stimare

Invece, nel caso di maggiore rilievo sociale, quello in cui le azioni, gli interventi e le politiche riguardino un numero ampio di destinatari, possiamo usare alcune delle tecniche che sono state sviluppate negli ultimi cento anni (e molte negli ultimi trenta) per stimare la situazione controfattuale e dunque l'effetto. Anche in questo caso, tuttavia, non saremo in grado di stimare effetti degli interventi su singoli destina-

ri, ma dovremo limitarci a misurare effetti “medi” su intere popolazioni. Si tratta certamente di una limitazione rilevante, ma una candela è meglio del buio totale.

Le tecniche sviluppate a partire dagli anni Trenta del secolo scorso, e soprattutto nel secondo dopoguerra, si possono suddividere in due grandi filoni. Da una parte ci sono le tecniche di tipo sperimentale, come l'esperimento randomizzato controllato (*Randomized Controlled Trial* – RCT) e dall'altra le tecniche non sperimentali, che fanno uso di strumenti statistici ed econometrici.

II RCT

L'elemento fondamentale del RCT è il processo di randomizzazione mediante il quale i soggetti potenzialmente destinatari dell'intervento o della politica vengono divisi (rigorosamente a caso) in due gruppi, normalmente denominati gruppo di trattamento e gruppo di controllo. Il gruppo di trattamento diventa il destinatario dell'intervento oggetto di valutazione, mentre il gruppo di controllo è normalmente escluso dallo stesso o, se possibile, riceve un placebo. Semplificando un po', possiamo dire che, quando il gruppo dei potenziali destinatari dell'intervento è sufficientemente ampio, il meccanismo di scelta causale e la legge statistica dei grandi numeri ci garantiscono che il gruppo di trattamento e quello di controllo siano – in media – identici. Per questa ragione, la situazione in cui si trova il gruppo di controllo, dopo che l'intervento è stato realizzato interessando il solo gruppo di trattamento, fornisce una buona stima della condizione in cui si sarebbe trovato il gruppo di trattamento se non fosse stato trattato, cioè della condizione controfattuale.

Per questo, l'effetto dell'intervento viene stimato sottraendo al valore medio della variabile di *outcome* misurata per il gruppo di trattamento (la situazione fattuale) il valore medio della stessa variabile misurata per il gruppo di controllo (la stima della situazione controfattuale). La Figura 1 illustra sinteticamente il processo descritto. Il RCT, se correttamente utilizzato, produce stime non distorte dell'effetto dell'intervento e rappresenta la miglior tecnica di cui disponiamo per stimare gli effetti di azioni, interventi e politiche.

Ovviamente, questa tecnica presuppone di essere pensata e attivata prima della realizzazione dell'intervento e dunque richiede di coinvolgere i valutatori sin dal momento in cui l'intervento stesso, o la politica, vengono progettati e implementati. Infatti, per poter assegnare a caso l'appartenenza al

gruppo di trattamento e a quello di controllo (evitando dunque il *selection bias*), bisogna che l'intervento non sia ancora stato avviato. Si tratta perciò di una tecnica non utilizzabile nel caso in cui l'esigenza valutativa sorga una volta che l'intervento sia già stato avviato o addirittura concluso.

Il coinvolgimento precoce dei valutatori potrebbe apparire come un limite di questa tecnica ma, in realtà, esso può invece rivelarsi uno dei suoi punti di forza. La presenza di valutatori esperti aiuta infatti i titolari degli interventi a dare una risposta chiara alle domande che è opportuno porsi prima di avviare ogni azione valutativa (e in realtà ogni intervento): vogliamo stimare l'effetto di che cosa (di quale intervento, definito in maniera precisa)? Su che cosa (su quali variabili di *outcome* effettivamente misurabili)? Aiutare i responsabili dei progetti a rispondere a queste domande è parte integrante del lavoro di un buon valutatore perché, senza risposte chiare, la stima degli effetti si rivela impossibile.

I RCT, pur non esenti da problemi, rappresentano un importantissimo strumento per non fidarci semplicemente delle sensazioni (o, peggio, dell'opinione di chi realizza l'attività) quando vogliamo capire se un intervento sociale o educativo sia efficace oppure no. Spesso i RCT sono oggetto di un'obiezione “etica”: per realizzarli dobbiamo negare l'intervento a qualche soggetto e ciò potrebbe creare un'iniquità. Nella gran parte dei casi, l'obiezione è priva di fondamento poiché l'esigenza di realizzare un RCT nasce proprio dalla nostra ignoranza riguardo all'efficacia di un intervento. Se l'intervento si rivelasse dannoso (esito non privo di plausibilità), l'iniquità riguarderebbe coloro che hanno ricevuto l'intervento, non coloro a cui lo stesso è stato negato. Se vogliamo capire che cosa funziona e che cosa no, spesso non possiamo fare altro che escludere qualcuno.

Ciò porta ad una ulteriore considerazione. Non è obbligatorio stimare gli effetti di un intervento. Poiché la stima è un'attività costosa, dovremmo compierla solo nei casi di genuina incertezza rispetto agli esiti della nostra azione. Se siamo sicuri che l'azione abbia effetti positivi (non importa come abbiamo acquisito questa certezza) la stima rappresenta solo tempo e risorse sottratte ad obiettivi di maggiore utilità sociale. Ma se non siamo certi degli esiti positivi, stimare con rigore è una esigenza imprescindibile, quasi un imperativo etico.

Va poi rilevato che talvolta i RCT sono ritenuti un metodo costoso per stimare gli effetti. Anche in questo caso, è necessario sfatare dei miti. Innanzitutto dobbiamo sottolineare che il costo della realizzazione dell'intervento (il doposcuola o altro) inciderà in egual misura qualunque sia la tecnica di stima dell'effetto che si intende utilizzare. Dunque le differenze di

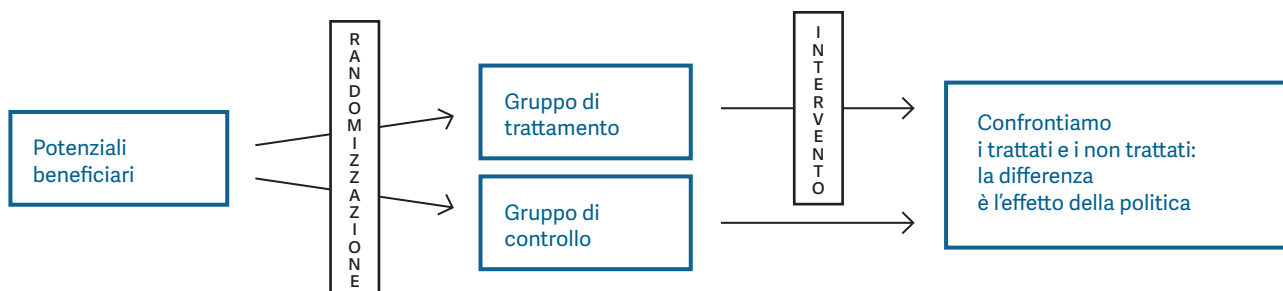


Figura 1
Randomized Controlled Trial – RCT

costo potrebbero dipendere da diversi oneri di raccolta delle informazioni necessarie ad effettuare le stime. Certamente la raccolta di informazioni sui soggetti trattati e sui controlli (necessaria in un RCT) è più costosa della raccolta di informazioni sui soli soggetti trattati. Ma quest'ultima è inutile (se non fuorviante) perché ci consente solo di ottenere stime distorte, mentre potremo fidarci maggiormente delle stime prodotte da un RCT ben realizzato. Inoltre, proprio la randomizzazione ci consentirà di raccogliere meno informazioni (più mirate) sui soggetti del gruppo di controllo di quanto dovremmo fare per produrre stime non distorte utilizzando le tecniche non sperimentali che illustreremo in un successivo paragrafo.

— Gli RCT e le politiche sociali ed educative

Esiste ormai una diffusa tradizione internazionale di utilizzo del RCT per stimare gli effetti di interventi educativi e sociali, anche se l'Italia fa un po' eccezione. Ad esempio, in campo educativo è importante ricordare l'esperienza di due soggetti, la *Education Endowment Foundations (EEF)*³ in Inghilterra e l'*Institute of Educational Sciences (IES)*⁴ negli Stati Uniti. La prima è una fondazione filantropica che vuole aiutare scuole e insegnanti a migliorare i risultati scolastici dei soggetti più svantaggiati, così da favorire l'accesso ai livelli superiori di istruzione anche a studenti di modeste condizioni familiari. La fondazione si dedica in particolare alla creazione e alla diffusione di conoscenza su ciò che è efficace nel migliorare l'apprendimento scolastico e per questo sostiene un gran numero di sperimentazioni, ne valuta gli effetti attraverso RCT e favorisce la diffusione delle pratiche efficaci attraverso un sito che descrive i diversi protocolli educativi e, per ciascuno di essi, riporta costi, stima dell'efficacia e robustezza dell'evidenza disponibile. In questo modo scopriamo, ad esempio, che gli interventi sull'architettura scolastica (spazi, banchi, disposizioni, ecc.), se valutati con rigore, mostrano una scarsa efficacia nel migliorare l'apprendimento dei ragazzi, e che le bocciature sono addirittura nocive. Al contrario, le pratiche di *collaborative learning* si rivelano utili e la didattica svolta durante le vacanze è molto efficace nel consentire di recuperare ritardi formativi. Poca ideologia, molta riflessione sulle esperienze e misurazione rigorosa dei risultati.

Lo IES è invece il braccio statistico e valutativo del Ministero dell'educazione statunitense. Anche questo organismo è votato a favorire lo sviluppo e la valutazione rigorosa di nuovi approcci che mirino a migliorare i risultati di apprendimento degli studenti. L'esito dell'attività di valutazione svolta attraverso gli ormai numerosissimi RCT sostenuti dall'istituto è riportato nella *What Works Clearinghouse*, un deposito preziosissimo di metodi e conoscenze a disposizione di operatori, insegnanti e policy makers. Per entrambe le istituzioni, i risultati degli RCT costituiscono la base empirica per la riforma (*evidence based*, come si dice oggi) dei sistemi educativi, oltre che per la modifica delle pratiche concrete degli insegnanti. Nel campo degli interventi sociali, vanno invece ricordate le

esperienze internazionali del *laboratorio J-Pal*⁵ e della *Campbell Collaboration*⁶. Il primo promuove studi randomizzati per produrre evidenza robusta sugli interventi che mirano a ridurre la povertà e a promuovere lo sviluppo. La sua rete di affiliati ha realizzato ormai oltre mille sperimentazioni sul campo, producendo informazioni e sostenendo la diffusione a popolazioni più ampie delle pratiche che localmente hanno mostrato esiti promettenti. I suoi direttori sono Abhijit Banerjee ed Esther Duflo, recenti vincitori del premio Nobel per l'economia proprio per i contributi metodologici ed empirici che hanno dato alle politiche di sviluppo.

La Campbell Collaboration realizza e pubblica *systematic reviews* (cioè analisi sistematiche delle valutazioni robuste) per studiare l'efficacia di alcuni interventi sociali in diversi settori del welfare. Il suo sito molto ricco riporta gli esiti di queste ampie analisi della letteratura riferita a specifici interventi sociali, mostrando che "non è tutto oro quel che luccica".

In tutte queste esperienze, il tentativo è quello di produrre conoscenza basata su metodi robusti (principalmente sugli esperimenti randomizzati) per mettere a disposizione degli operatori e dei policy makers strumenti che consentano loro di migliorare la propria attività, evitando di basarsi su cosiddette *best practice* che spesso sono tali solo nella mente di coloro che le promuovono, senza essere mai state oggetto di valutazioni rigorose.

— Quando non si possono fare RCT

Nei casi in cui non sia possibile realizzare un RCT – perché l'intervento è già stato avviato o perché, per ragioni tecniche o politiche, non è immaginabile di escludere qualche utente dall'intervento stesso – non si disporrà automaticamente di un gruppo di controllo che sia equivalente (in media) al gruppo dei soggetti trattati, così da poterlo utilizzare per la stima della situazione controfattuale. Per queste ragioni, la situazione controfattuale viene stimata utilizzando dati relativi a gruppi di soggetti non interessati dall'intervento, utilizzando tecniche di tipo statistico ed econometrico⁷ che provano a risolvere i due problemi fondamentali della stima d'effetto, cioè la distorsione da selezione e quella da dinamica spontanea. A titolo esemplificativo, e non esaustivo, si possono citare le tecniche dell'abbinamento statistico e della differenza-nelle-differenze.

Con la *tecnica dell'abbinamento statistico* si prova a tenere sotto controllo la distorsione da selezione. L'intuizione di fondo è che, per stimare la situazione controfattuale, si debbano trovare, all'interno del gruppo dei soggetti che non hanno partecipato all'intervento, gli individui più simili a quelli che hanno partecipato allo stesso. In questo modo, confrontando soggetti simili, si può immaginare che le eventuali differenze riscontrate nelle variabili di *outcome* possano essere attribuite all'intervento.

Sono state proposte diverse tecniche statistiche che aiutano ad individuare i "casi gemelli". Tutte le tecniche scontano però un

3 <https://educationendowmentfoundation.org.uk>

4 <https://ies.ed.gov/>

5 <https://www.povertyactionlab.org/>

6 <https://campbellcollaboration.org/>

7 In questa sede non è possibile descrivere in dettaglio le tecniche non sperimentali utilizzate per la valutazione. Per una descrizione

approfondita di queste tecniche si vedano Martini e Sisti (2009) o Gertler et al. (2016).

problema comune: all'interno dei due gruppi, potremo individuare soggetti simili solo sulla base di caratteristiche osservabili (ad esempio il genere, l'età e così via), mentre l'abbinamento sulla base delle caratteristiche non osservabili (ad esempio la motivazione, la determinazione, l'altruismo, le capacità relazionali, ecc.) è impossibile da realizzare. Pertanto, dal punto di vista delle variabili non osservabili, non si può escludere l'esistenza di fenomeni di distorsione da selezione quando si confrontano i due gruppi. Ciò non è privo di conseguenze perché, specie nel caso di interventi sociali ed educativi, proprio le caratteristiche non osservabili degli individui giocano spesso un ruolo chiave nell'influenzare le variabili di *outcome*.

Pertanto, l'utilizzo delle tecniche di abbinamento non garantisce di poter ottenere stime non distorte dell'effetto degli interventi; tanto più numerose saranno le caratteristiche che siamo in grado di osservare, tanto migliore potrà essere l'abbinamento, anche se qualcosa resterà sempre impossibile da osservare e misurare. Pertanto, queste tecniche sono particolarmente esigenti dal punto di vista delle informazioni che debbono essere disponibili e misurabili, non solo per coloro che prendono parte all'intervento, ma anche per i soggetti che vengono utilizzati come gruppo di controllo.

Con la tecnica della *differenza-nelle-differenze* si prova invece a tenere sotto controllo la distorsione da dinamica spontanea. L'intuizione di fondo è che la differenza tra la misura post e la misura pre-intervento della variabile di *outcome* per i soggetti trattati sia influenzata tanto dall'effetto dell'intervento quanto dalla dinamica spontanea. Pertanto, per correggere la distorsione, sarebbe necessario conoscere in che modo la dinamica spontanea abbia contribuito al cambiamento della variabile di *outcome*. La tecnica della *differenza-nelle-differenze* ipotizza che la dinamica spontanea della variabile di *outcome* per il gruppo di trattamento (che non possiamo misurare, perché non siamo in grado di separarla dall'effetto dell'intervento stesso) sia identica a quella misurata per il gruppo di controllo, che invece possiamo misurare tranquillamente poiché questo gruppo non è interessato dall'intervento (si chiama gergalmente "ipotesi di parallelismo"). Di conseguenza, per stimare l'effetto dell'intervento si calcolano tre differenze: a) la differenza tra il valore della variabile di *outcome* del gruppo dei trattati dopo l'intervento e quello prima dell'intervento; b) la stessa differenza nel gruppo dei soggetti non trattati e c) la differenza tra i due valori precedenti, che rappresenta la stima dell'effetto.

In questo caso, la presenza di molte informazioni sull'andamento della variabile di *outcome* nel periodo che precede l'intervento rende più credibile l'ipotesi di parallelismo. Quindi, come l'abbinamento statistico, anche questa tecnica è piuttosto esigente dal punto di vista dei dati.

Ovviamente le due tecniche possono essere usate congiuntamente quando immaginiamo che entrambi i tipi di distorsione siano rilevanti per stimare l'effetto dell'intervento che stiamo realizzando.

Conclusioni

Le brevi riflessioni esposte ci consentono di proporre alcune conclusioni sintetiche. Il nostro sistema di welfare ha un forte bisogno di innovazioni che provino a rispondere a bisogni emergenti o a dare risposte nuove a bisogni antichi ma irrisolti. Le innovazioni sociali sono intrinsecamente rischiose poiché potrebbero non funzionare o addirittura rivelarsi nocive. Per questo è indispensabile utilizzare strumenti affidabili e precisi per stimare gli effetti di queste innovazioni, così da consentire agli operatori e ai policy makers di prendere decisioni informate, basate su "evidenze robuste" e non su semplici sensazioni. Una maggiore attenzione a queste evidenze potrebbe consentire sia di ridurre il costo dei servizi sia di aumentarne l'efficacia, a tutto beneficio dei cittadini e degli utenti.

La buona notizia è che queste "evidenze robuste" non sono impossibili da produrre, al contrario, molti altri paesi già si sono impegnati in questa direzione. Serve una consapevolezza chiara dei benefici che queste informazioni potrebbero produrre per la collettività e la decisione collettiva (di un insieme di organizzazioni del Terzo settore? Dei policy makers a diversi livelli?) di destinare risorse (umane ed economiche) alla loro produzione. È necessario uscire dal paradosso che ci vede dedicare molta attenzione a obiettivi irraggiungibili, come la stima dell'impatto di una organizzazione, e trascurare completamente obiettivi raggiungibili e estremamente utili, come la stima degli effetti di interventi e politiche che interessano ampia parte della nostra popolazione.

DOI: 10.7425/IS.2020.04.13

Bibliografia

Conti G., Heckman J., Pinto R. (2016), "The Effects of Two Influential Early Childhood Interventions on Health and Healthy Behaviour", *The Economic Journal*, 126(596), pp. 28-65.

Gertler P.J., Martinez S., Premand P., Rawlings L.B., Vermeersch C.M.J. (2016), *Impact Evaluation in Practice*, Second Edition, Inter-American Development Bank and World Bank, Washington DC.

Holland P. (1986), "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81(396), pp. 945-960.

Martini A., Sisti M. (2006), *Valutare il successo delle politiche pubbliche*, Il Mulino, Bologna.