

SOSE WEST

Workshop Economico, Statistico e Tecnologico

DATA SCIENCE

PER LE POLITICHE

PUBBLICHE

Roma - 30 OTTOBRE 2019

**LA MISURAZIONE DEGLI EFFETTI CAUSALI DI
UN INTERVENTO: LOGICA E METODI
PRINCIPALI.**

*Enrico Rettore,
Università di Trento e FBK-
IRVAPP*

1. INTRODUZIONE

- Un intervento produce degli effetti se *fa* capitare qualcosa. Cioè, se quel qualcosa *non* capiterebbe in assenza dell'intervento stesso.
- Ad es., *corsi di formazione*: hanno un effetto sul tasso di occupazione dei partecipanti se in assenza dei corsi gli *stessi* soggetti - negli *stessi* mesi - avrebbero sperimentato un tasso di occupazione diverso da quello osservato.
- *Effetto causale* di un intervento: definito come differenza tra ciò che succede ai soggetti esposti - il loro risultato *fattuale* - e ciò che gli succederebbe - se non venissero esposti all'intervento - il loro risultato *controfattuale*.

- Dalla definizione alla misurazione. Ad intervento svolto, il risultato *fattuale* è osservabile, il risultato *controfattuale* è irreversibilmente perso.
- Unica soluzione praticabile: approssimare il risultato controfattuale dei soggetti esposti con il risultato fattuale di un *opportuno* gruppo di soggetti non esposti.
---> *gruppo di confronto.*
- *Impossibile* verificare se questo gruppo di non esposti approssima in modo accurato il risultato controfattuale degli esposti.
- Tutte le strategie di misurazione dell'impatto poggiano su *un'ipotesi non verificabile.*

- Come si deve procedere *in pratica* per selezionare l'opportuno gruppo di confronto?
- Questo è il passaggio cruciale di qualsiasi valutazione dell'impatto di un intervento.
- Più in generale, è il problema che si pone *sempre* quando si vogliono identificare relazioni di causa ed effetto.
- Tensione tra *validità interna* e *validità esterna*.

2. IL *RANDOMIZED CONTROL TRIAL* (RCT)

- In un RCT la scelta dei soggetti da esporre/non esporre all'intervento avviene in modo casuale: l'equivalente di un lancio di moneta.
- Ne risultano due gruppi *mediamente equivalenti* in ogni aspetto.
- Conseguenza fondamentale: l'esito *fattuale* - osservato - dei soggetti non esposti è mediamente eguale all'esito *controfattuale* - non osservato - dei soggetti esposti.

---> Le differenze tra i due gruppi nei loro esiti sono interpretabili *causalmente*.

- In linea di principio, è il migliore modo per approssimare il risultato controfattuale dei soggetti esposti all'intervento.
- Obiezione etica: la lotteria esclude ingiustamente dall'intervento soggetti che potrebbero beneficiarne.
- Obiezione infondata almeno nel caso degli *interventi pilota*: messi in atto per scoprire *se* l'intervento serve...

- Problemi operativi 1: *non compliance*.

Soggetti assegnati all'esposizione all'intervento che si sottraggono; soggetti assegnati al gruppo di controllo che trovano il modo di essere esposti all'intervento.

- Più in generale: eventi successivi alla randomizzazione che alterano il protocollo sperimentale. Inevitabile quando si ha a che fare con soggetti umani.

- Problemi operativi 2: validità esterna.

- RCT utilizzati per valutare l'impatto di innumerevoli interventi pubblici, in paesi sviluppati e in via di sviluppo (Duflo e Banerjee, 2017).

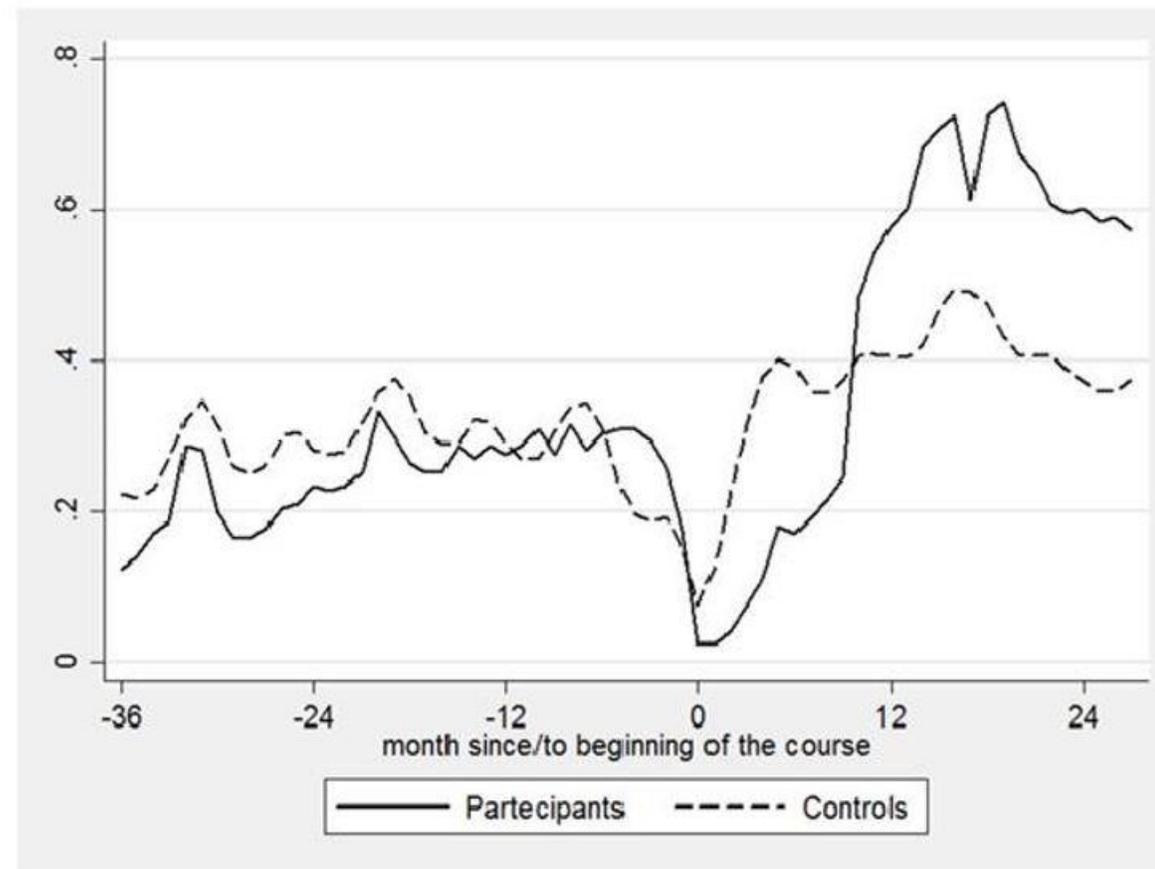
- Cosa si può fare se non è possibile condurre un RCT?
- Differenza rispetto al caso RCT: il valutatore non ha alcun controllo sul processo di selezione (o lo ha solo parzialmente). Può solo *osservare* (o parzialmente influire su) ciò che accade attorno a lui.
---> Disegni osservazionali
- Obiettivo invariato: selezionare un gruppo di soggetti non esposti all'intervento che consentano di approssimare accuratamente il risultato controfattuale dei soggetti esposti.
- Quattro casi per esemplificare.

3. I DISEGNI OSSERVAZIONALI 1: IL CONTROLLO DELLE DIFFERENZE OSSERVABILI TRA ESPOSTI E NON ESPOSTI

- Corsi di formazione, Provincia Autonoma di Trento, 2010 - 2011
- I soggetti ammissibili *scelgono* liberamente se partecipare o meno. Cioè, il valutatore *non* ha la minima possibilità di influire sul processo di selezione.
- Nel caso considerato, il gruppo di confronto è costituito da soggetti che hanno avuto l'opportunità di partecipare ad uno dei corsi e che hanno *scelto* di non partecipare
- Pari età, livello di istruzione, area geografica, condizione occupazionale alla data di inizio dei corsi tra partecipanti e non partecipanti.

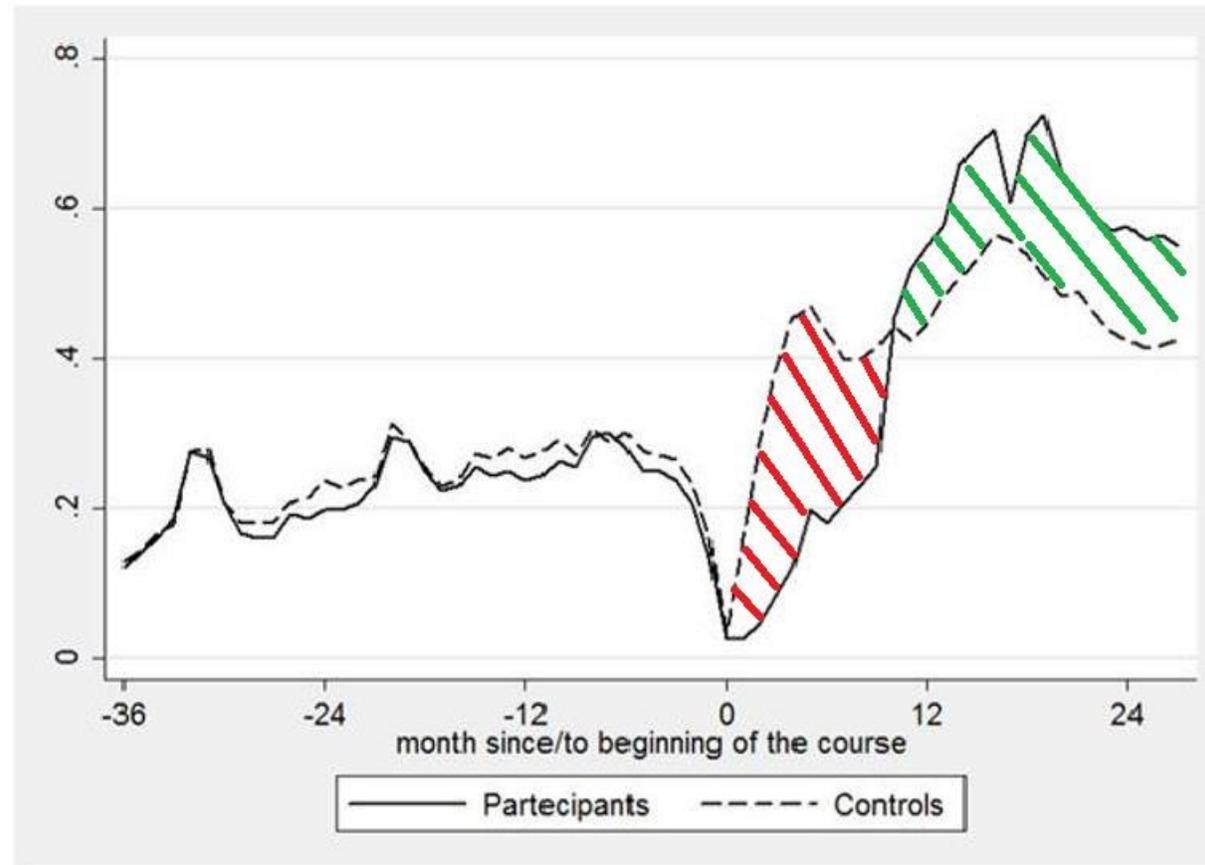
- I due gruppi sono *comparabili*?
- Se lo fossero, nei mesi precedenti l'inizio dei corsi - mesi nei quali l'effetto dei corsi *non* può ancora manifestarsi - i due gruppi dovrebbero presentare storie lavorative comparabili.
- Chiari segni che i partecipanti ai corsi presentano tassi di occupazione *sistematicamente inferiori* ai non partecipanti, nell'arco dei tre anni precedenti l'inizio dei corsi.

Figura 3 – Tassi di occupazione dei partecipanti e dei non partecipanti ai corsi di formazione da 36 mesi prima a 27 mesi dopo l’inizio dei corsi stessi, Provincia Autonoma di Trento 2010 e 2011.



- Soluzione: *mimare* ex post ciò che si farebbe se fosse possibile condurre un RCT.
- Il gruppo di confronto viene selezionato scegliendo tra i non partecipanti coloro che risultano più simili ai partecipanti rispetto alle *informazioni disponibili*.
- Nei 36 mesi precedenti l'inizio dei corsi i tassi di occupazione dei due gruppi sono - per costruzione! - pressoché identici: gruppo di confronto selezionato includendovi *solo* i soggetti con carriere occupazionali comparabili a quelle dei partecipanti.
- Effetto *lock-in* fino al 10° mese dall'inizio del corso. Poi effetto positivo, 10-15 p.p.

Figura 4 – Tassi di occupazione dei partecipanti e dei non partecipanti ai corsi di formazione, abbinati mediante matching, da 36 mesi prima a 27 mesi dopo l’inizio dei corsi stessi, Provincia Autonoma di Trento 2010 e 2011.



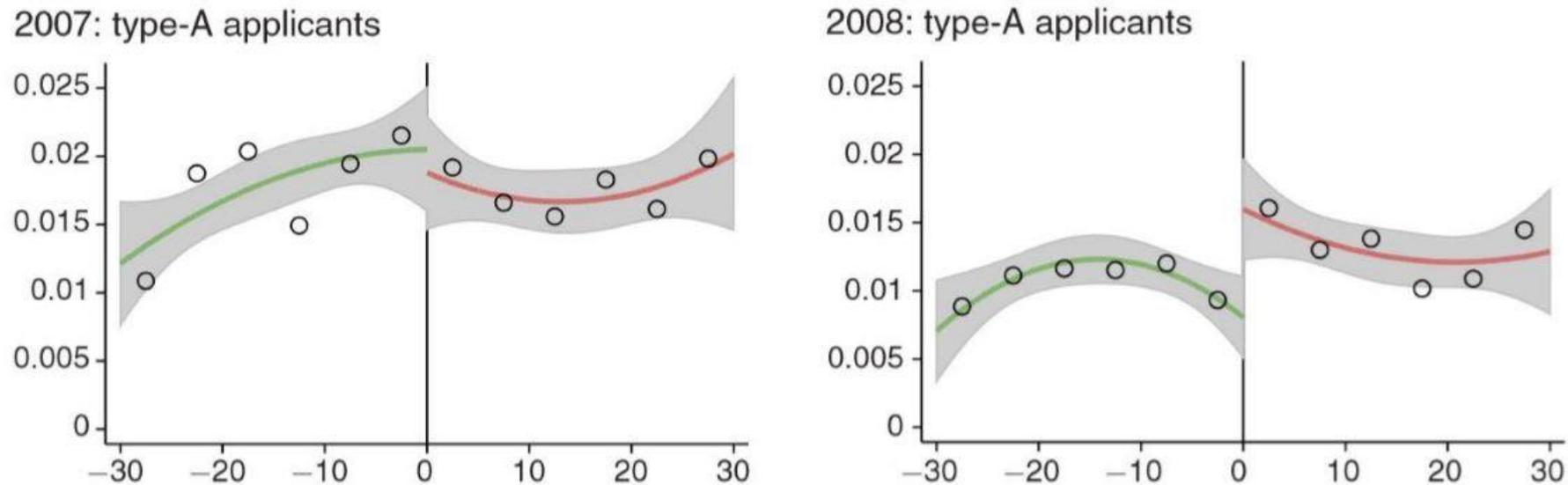
- La soluzione che *mima ex-post* un RCT dà luogo a due gruppi equivalenti in media rispetto alle sole caratteristiche rispetto alle quali i due gruppi vengono forzati ad essere equivalenti.
- Differenze rispetto a caratteristiche non osservabili?
- Ruolo di archivi amministrativi 'ricchi' ---> esposti e non esposti resi comparabili rispetto a ampio insieme di caratteristiche osservabili.
- Problemi tecnici se le caratteristiche rispetto alle quali stabilire l'equivalenza sono 'troppe' ---> *machine learning methods* (Mullainathan e Spiess, 2017).

4. I DISEGNI OSSERVAZIONALI 2: IL *REGRESSION DISCONTINUITY DESIGN* (RDD)

- Effetti della regolarizzazione degli immigrati sul loro tasso di criminalità (Pinotti, 2017).
- I dati si riferiscono a coloro che hanno presentato domanda di regolarizzazione a dicembre 2007.
- Dalle ore 8.00 del giorno stabilito, i datori di lavoro/sponsor dei richiedenti la regolarizzazione hanno presentato la domanda on-line. Regolarizzazioni concesse secondo l'ordine di presentazione delle domande, fino all'esaurimento della *quota* stabilita.

- Selezione dei regolarizzati/non regolarizzati basato su una *regola sistematica* - il momento della presentazione della domanda.
 ---> I 'ritardatari' sono comparabili ai 'rapidi'?
- Una buona soluzione consiste nel restringere l'analisi al sottoinsieme dei soggetti che hanno ottenuto/mancato, la regolarizzazione di poco.
- I primi esclusi sono molto simili agli ultimi ammessi, rispetto all'*unica* caratteristica rilevante per la regolarizzazione: il momento della presentazione della domanda.
 ---> nell'anno *precedente* la presentazione della domanda (2007) *non* si osserva alcuna differenza nel tasso di criminalità tra i *due gruppi al margine*.

Figura 5 – Numero medio di crimini commessi per richiedente nell’anno precedente (a sinistra) e nell’anno successivo (a destra) la presentazione della domanda secondo l’esito della domanda: in verde i regolarizzati, in rosso i non regolarizzati. L’asse orizzontale misura il tempo in minuti dal momento della domanda dell’ultimo ammesso.



- Nell'anno successivo (2008) la regolarizzazione *dimezza* il tasso di criminalità (benchmark 15 per mille).
- L'effetto della regolarizzazione misurato in questo modo si riferisce ai soli *ammessi al margine*.
- Questa strategia consente di ottenere una stima credibile dell'effetto dell'intervento al prezzo di restringere l'analisi ad un particolare sottoinsieme di soggetti.
---> validità interna solida, valida esterna debole (in generale)
- Sviluppi recenti per generalizzare (Angrist, Rokkanen, 2015).

5. I DISEGNI OSSERVAZIONALI 3: LA DIFFERENZA DI DIFFERENZE (DIFF-IN-DIFF'S).

- Il classico lavoro di Card e Krueger (1995) sugli effetti occupazionali del salario minimo.
- Il salario orario minimo in New Jersey (NJ) *aumenta* da \$4.25 a \$5.05, a partire da Apr. 1992. *Non* cambia in Pennsylvania (PA): \$4.25.
- Campione di ristoranti fast-food, intervistati due volte: Feb. e Dic. 1992.
- Nonostante la prossimità geografica, ci sono differenze tra i ristoranti NJ e PA, già a Feb. 1992.

Figura 6 – La mappa geografica del disegno di valutazione.

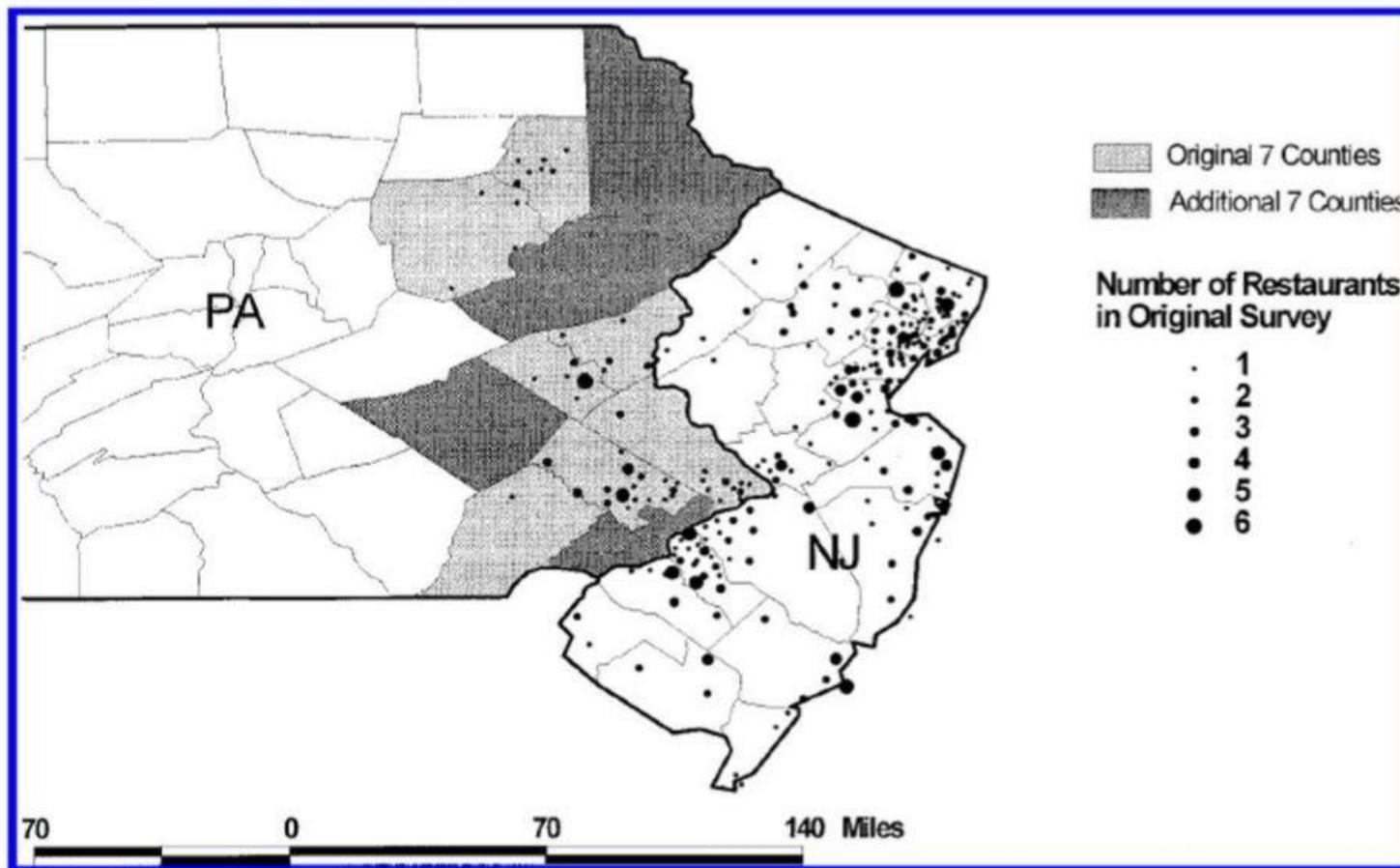


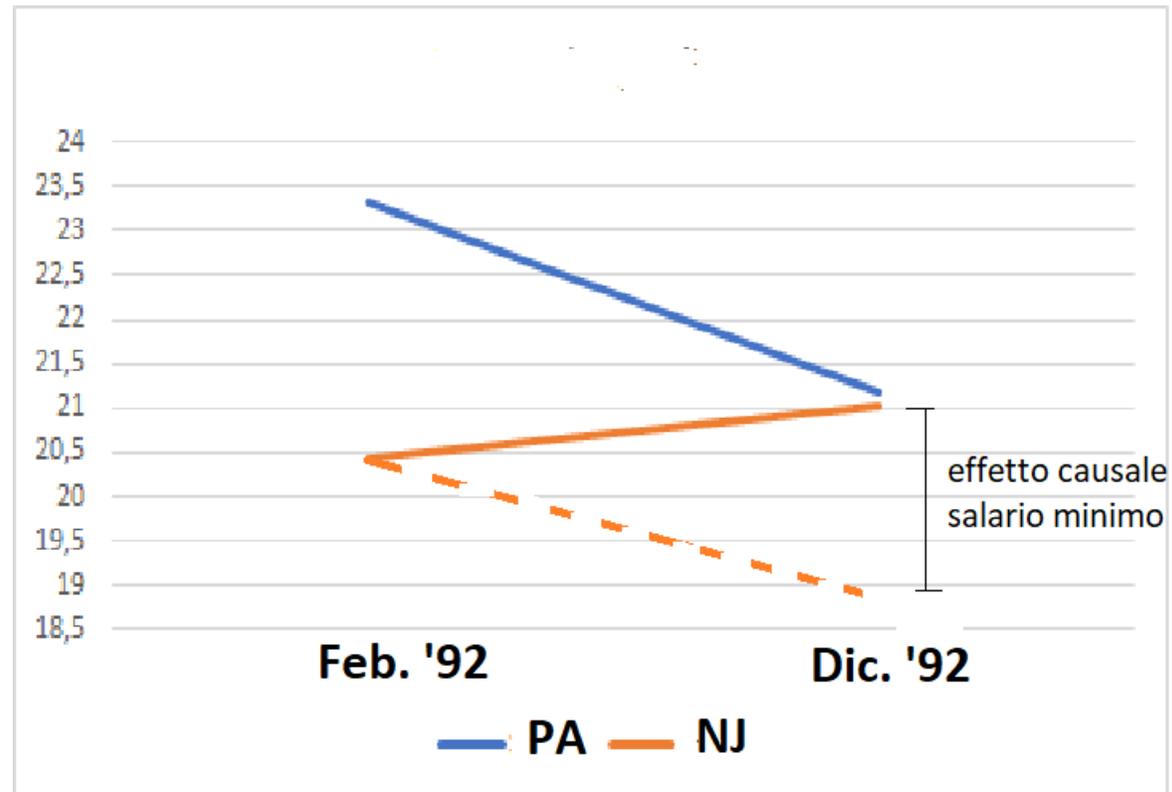
FIGURE 1. AREAS OF NEW JERSEY AND PENNSYLVANIA COVERED BY ORIGINAL SURVEY AND BLS DATA

- Questa evidenza è sufficiente per dubitare che i ristoranti PA approssimino in modo accurato il risultato controfattuale dei ristoranti NJ a Dic. 1992.
- La soluzione adottata. Ipotesi di *trend comune*: in assenza della riforma, la *variazione dell'occupazione* tra Feb e Dic 1992 per i ristoranti NJ *sarebbe* stata identica a quella osservata per i ristoranti PA

$$\text{variazione } \textit{fattuale} \text{ PA} = \text{variazione } \textit{controfattuale} \text{ NJ}$$

- Se vale questa ipotesi, l'effetto cercato è pari a:
variazione *fattuale* dell'occupazione in NJ - variazione *fattuale* occupazione in PA

EFFETTO DELL'AUMENTO DEL SALARIO MINIMO IN NJ, APR. '92. LINEA CONTINUA=TREND FATTUALE, LINEA TRATTEGGIATA=TREND CONTROFATTUALE



6. I DISEGNI OSSERVAZIONALI 4: LE VARIABILI STRUMENTALI

- Il caso considerato esemplifica i problemi operativi di RCT.
- Il valutatore *sceglie* chi deve essere esposto all'intervento. Ma i suoi sforzi per tenere sotto controllo il processo di selezione sono in parte vanificati dal comportamento sul campo dei vari attori coinvolti.
- Uno studio nato per essere RCT finisce per essere in buona misura osservazionale.
- *Lavoro&Psiche* è un *intervento pilota* messo in atto per studiare gli effetti di una

- 311 soggetti, bipartiti casualmente tra gruppo di trattamento e gruppo di controllo.
- Periodo sperimentale: 2011 e 2012. I soggetti inclusi nel gruppo di trattamento sono stati seguiti da un *job coach*, interamente dedicato al supporto nella ricerca di lavoro di un piccolo numero di soggetti (non più di 12-13 per addetto).
- Effetti dell'intervento valutati con riferimento agli *esiti occupazionali* nel 2013.
- Analisi di implementazione: il principale ruolo svolto dal *job coach* è consistito nel facilitare l'accesso ad un tirocinio ai soggetti assistiti.

- L'effetto causale dell'intervento sugli esiti lavorativi - se c'è - passa dall'esperienza di un tirocinio.
---> effetto causale del *tirocinio* vs del *job coach*
- Punto critico per la valutazione: non tutti i soggetti inclusi nel gruppo di trattamento sperimentano un periodo di tirocinio; alcuni tra i soggetti inclusi nel gruppo di controllo accedono ad un tirocinio.
- Classico problema in un RCT: anche se l'assegnazione causale *fa* la differenza per l'effettiva esposizione all'intervento - il tirocinio - difficilmente la determina in modo univoco.

Figura 8 – Tassi di partecipazione ad un tirocinio per i soggetti inclusi, rispettivamente, nel gruppo di trattamento e nel gruppo di controllo.



- Si parte dal confronto tra i due gruppi selezionati mediante randomizzazione. In questo modo si identifica l'effetto di avere avuto a disposizione un *job coach*. In letteratura, *Intention to Treat* (ITT).
- Si riscalda l'ITT per tenere conto del fatto che:
 $\Pr(\text{tirocino}|\text{job coach})=0.55$
 $\Pr(\text{tirocino}|\text{no job coach})=0.25$
- Nel caso considerato:
Effetto medio tirocinio su probabilità di lavorare almeno un giorno = 0.19
Effetto medio tirocinio sul numero di giorni lavorati = 50

- Soluzione sfrutta i benefici della randomizzazione - l'equivalenza tra i due gruppi selezionati casualmente - e apporta una correzione ex post per tenere conto dell'alterazione del protocollo sperimentale.
- Il prezzo di questa soluzione. Imbens e Angrist (1994) hanno mostrato che in questo modo si ottiene un effetto medio su un particolare sottogruppo di soggetti - i *compliers*.
---> validità interna solida, valida esterna debole (in generale)
- Nel caso considerato: *compliers*=accedono al tirocinio se e solo se vengono assistiti da un *job coach*. Sono ca. il 30% del campione.

- L'effetto medio trovato in questo modo in generale *non* può essere generalizzato ai non *compliers*.
- Molta letteratura sulla generalizzazione (Imbens e Rubin, 1997...).
- Nel caso specifico, evidenza favorevole alla generalizzazione (Martini, Rettore e Barbetta, 2018).

7. CONSIDERAZIONI CONCLUSIVE

- La logica controfattuale è alla base del metodo scientifico di identificazione delle relazioni di causa ed effetto.
- Ad es., Diamond (1997): a causare le diseguaglianze nel grado di sviluppo delle varie parti del globo sono le loro caratteristiche geografiche, non differenze biologiche tra le varie popolazioni:

‘I expect that if the populations of Aboriginal Australia and Eurasia *could have been* interchanged during the Late Pleistocene, the original Aboriginal Australians *would now be* the ones occupying most of the Americas and Australia, as well as Eurasia, while the original Aboriginal Eurasians *would be* the ones now reduced to downtrodden population fragments in Australia.’

- Come approssimare il controfattuale?
- Se praticabile, RCT è senza dubbio il modo più affidabile per ottenere stime credibili dell'effetto causale di un intervento.
- Ma... serve essere preparati a procedere in modo osservazionale.
- Anche nel RCT meglio disegnato è facile che succeda qualcosa di non previsto dal valutatore - e fuori dal suo controllo - che in qualche misura altera il protocollo sperimentale, rendendo necessarie correzioni ex post.

- Ci sono miriadi di domande interessanti riguardanti relazioni causali nelle scienze sociali, molte meno opportunità di condurre RCT.
- Sarebbe un peccato non dare risposta a quelle domande solo perché non c'è modo di condurre un RCT.
- Più importante della distinzione RCT vs disegno osservazionale è la distinzione tra valutazioni *retrospettive* e *prospettive*:

V. prospettive: progettate *assieme* all'intervento da valutare

V. retrospettive: *improvvisate* ad intervento concluso

- La valutazione fallisce se non si riesce ad identificare un adeguato gruppo di confronto.
- Per ridurre il rischio di fallimento serve progettare la valutazione non appena si inizia a progettare l'intervento oggetto di valutazione.
- Rubin (2007): 'The design vs the analysis of observational studies for causal effects'

...[RCT] are often infeasible to conduct for a variety of reasons, such as ethical concerns, excessive expense, or timeliness. Consequently, much of our knowledge of causal effects must come from non-randomized observational studies...

...observational studies can and should be designed to approximate randomized

...observational studies should be designed ... to create subgroups of similar treated and control units, where 'similar' here refers to their distributions of background variables...

...this activity should be conducted without any access [*prior*] to any outcome data, thereby assuring the objectivity of the design.