

Metodi Statistici per l'Analisi Socio-Economica

Parte II

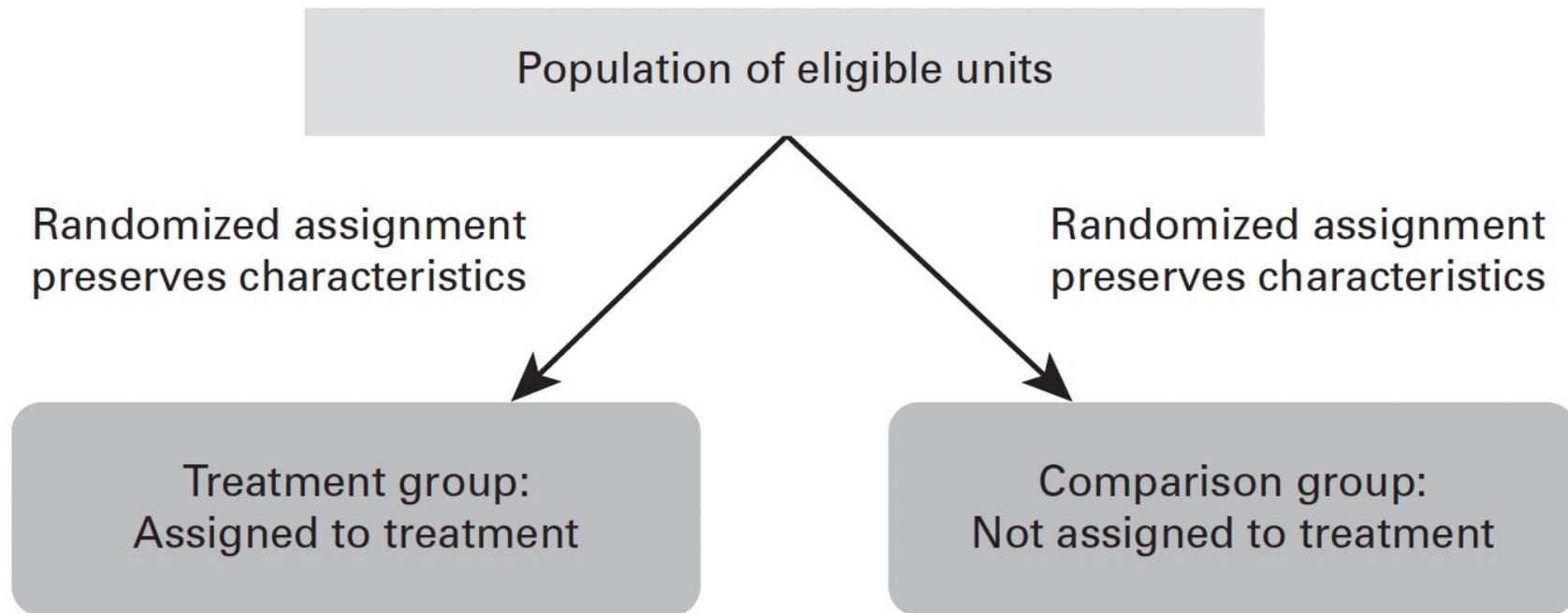
Introduzione alla valutazione di un intervento

Lezione 4



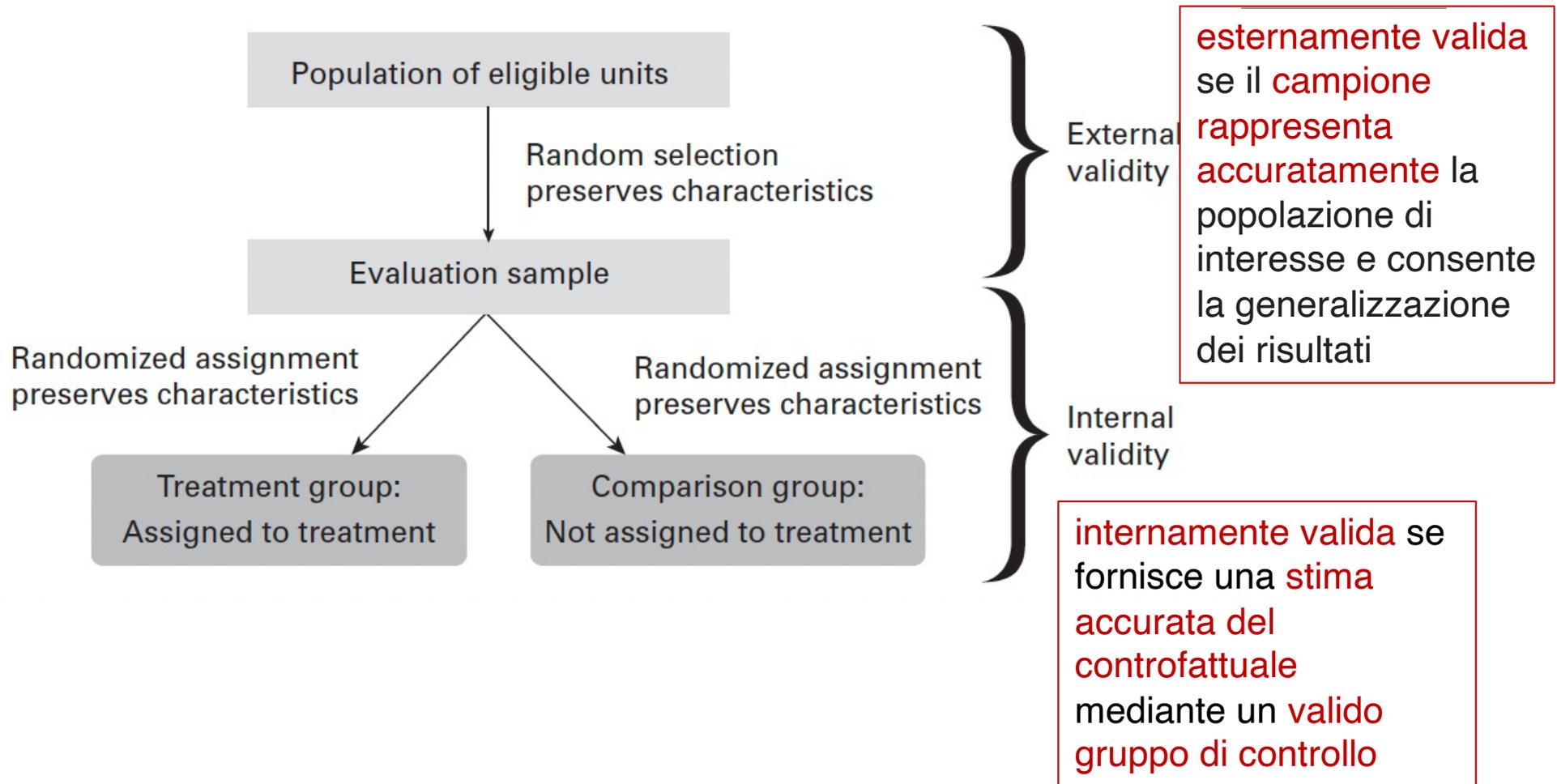
Randomizzazione e caratteristiche unità

Figure 4.1 Characteristics of Groups under Randomized Assignment of Treatment



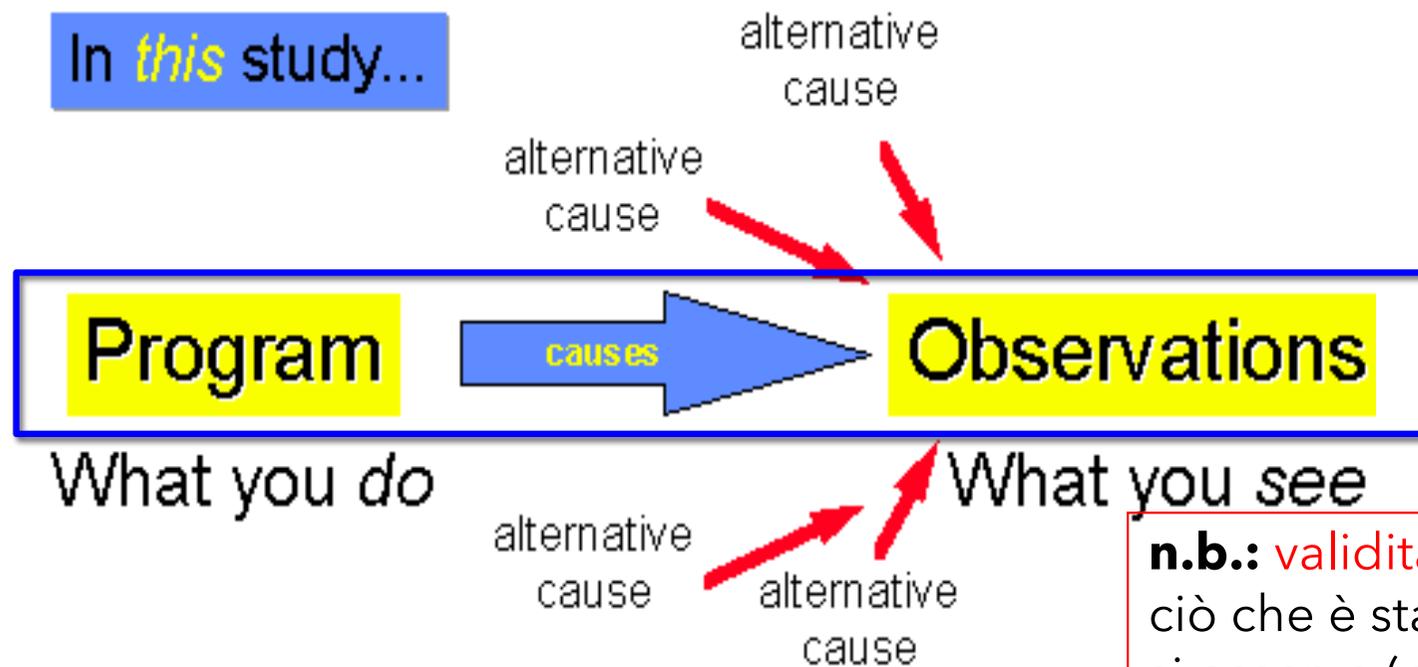
Valutazione d'impatto e validità interna ed esterna (con randomizzazione)

Figure 4.2 Random Sampling and Randomized Assignment of Treatment



Valutazione d'impatto e validità interna

Una valutazione d'impatto è detta avere **validità interna** se consente di individuare correttamente l'impatto della politica sulle unità coinvolte nello studio
(l'intervento ha prodotto un effetto)



n.b.: **validità interna** assicura che ciò che è stato fatto *causa* ciò che si osserva (gruppo di controllo è una buona approssimazione del controfattuale **come in randomizzazione**)

Valutazione d'impatto e validità interna

(in generale)

Minacce alla validità interna:

problemi nel disegno e/o nelle modalità di valutazione della politica che possono portare a stimare in maniera errata l'impatto della politica sugli esposti.

Fattori che possono minacciare la validità interna

selezione (non casuale)

storia (\cong dinamica spontanea)

contaminazione

Selezione

possibilità che ci siano delle differenze iniziali tra i due gruppi che influenzano la variabile risultato e che quindi confondano l'effetto del trattamento

(minaccia tipica del **confronto trattamento/controllo**)

Le cause possono essere:

- autoselezione
- etero-selezione

Storia

(\cong dinamica spontanea): possibilità che qualcosa d'altro rispetto al trattamento sia responsabile di tutto o di una parte del cambiamento osservato nel tempo.

(minaccia tipica del **confronto prima/dopo**, anche considerando "trattati" e "non trattati")

Le cause principali possono essere:

- eventi esterni intervenuti -in modo diverso-
- maturazione (processo di invecchiamento o sviluppo naturale) diversa
- effetto "pre-testing" (test iniziale a cui seguono reazioni diverse)
- "attrition": uscita dal gruppo di "trattati" e "non trattati" differenziata (sia come tasso di uscita che caratteristiche)

Contaminazione

possibilità che il gruppo di trattamento e/o di controllo venga **contaminato** (\equiv reso differente dallo stato "puro" ipotizzato) da qualche altro fattore nel corso dell'attuazione della politica, con conseguenze sugli effetti della stessa.

Le principali cause possono essere:

- diverso comportamento del gruppo di controllo dovuto alla presenza del trattamento (*control group substitution bias*): imitazione /competizione /demoralizzazione

e anche

- cattiva somministrazione del trattamento (es: dosaggi errati sui trattati)
- controindicazioni o effetti non previsti nella somministrazione che causano problemi nella variabile risposta

Disegno sperimentale e validità interna

selezione casuale = come viene estratto il campione dalla popolazione
(validità esterna)

assegnazione casuale = come il campione è assegnato ai trattamenti
(validità interna)

history	✓
maturation	✓
testing	✓
instrumentation	✓
mortality	✓
regression to the mean	✓
selection	✓
selection-history	✓
selection- maturation	✓
selection- testing	✓
selection- instrumentation	✓
selection- mortality	✗
selection- regression	✓
diffusion or imitation	✗
compensatory equalization	✗
compensatory rivalry	✗
resentful demoralization	✗

Valutazione d'impatto e validità esterna

Una valutazione d'impatto è detta avere **validità esterna** se i risultati ottenuti possono **essere estesi** ad una popolazione più vasta dei soli individui coinvolti nello studio.

Minacce alla validità esterna

Si riferiscono alla *generalizzabilità* dei risultati di una valutazione d'impatto

fattori che impediscono l'estensione ad altri soggetti (e, più in generale, altri periodi di tempo, altri *settings* – *demonstrations vs. politiche a regime* –, altre operazionalizzazioni del trattamento o altre caratteristiche della variabile obiettivo).

Uno studio di caso: National Supported Work Demonstration (USA, 1975-78)

- Popolazione obiettivo:
 - madri non sposate che da almeno 3 anni vivono del sussidio di povertà
 - giovani *drop-outs* della scuola dell'obbligo, con precedenti di piccola criminalità
 - ex tossicodipendenti
 - ex carcerati
- Obiettivo politica: inserimento nel mercato di lavoro regolare
- Contenuto della politica: inserimento dei soggetti in un'esperienza lavorativa reale ma in un ambiente "protetto" (aumento graduale dello sforzo, supervisione). Si prevede che il periodo di *supported work* duri circa 1 anno, ad una retribuzione poco superiore al salario minimo
- Costo: \$5.000-8.000 (valori del 1975) per partecipante

Perché demonstration

- per valutare l'efficacia della politica su scala ridotta prima di decidere sull'applicazione su larga scala
- Costo totale: \$82 milioni dei quali \$11 milioni per condurre la valutazione (spese a carico del governo federale, degli stati in cui la *demonstration* ha luogo e di alcune fondazioni)
- Luogo di realizzazione: 14 centri urbani in altrettanti stati (Atlanta, Detroit, Chicago, New York, Philadelphia, San Francisco, etc.)
- Obiettivi della valutazione: rispondere ai quesiti
 - il *supported work* aumenta la partecipazione al lavoro degli esposti e diminuisce la loro dipendenza dai sussidi pubblici?
 - quali gruppi beneficiano maggiormente della politica?
 - quali sono i costi? E il rapporto costi-benefici?
 - quali strategie di implementazione funzionano meglio?
 - quali caratteristiche della politica hanno maggiore impatto sui partecipanti?

Alcune considerazioni

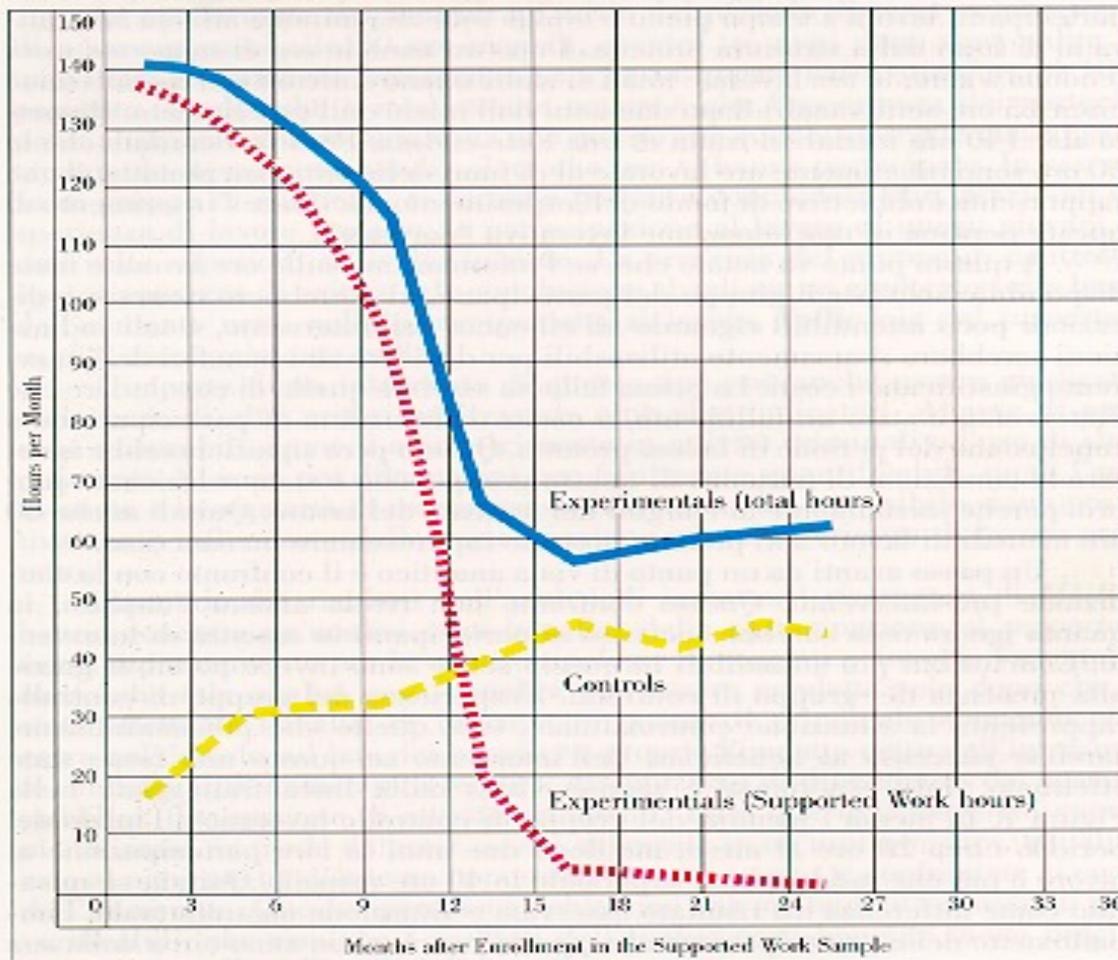
- NB1. Non vi sono ostacoli di ordine etico alla randomizzazione: se non vi fosse la volontà esplicita di condurre l'esperimento la politica (su scala ridotta) non avrebbe luogo.
- NB2. Vi sono comunque resistenze da parte degli operatori sul campo incaricati di selezionare $2n$ unità e poi di annunciare a n di queste che non avrebbero preso parte al programma.
- NB3. Qualcuno potrebbe obiettare che, fermo restando il vincolo di bilancio, e quindi la necessità di negare a qualche richiedente l'accesso al programma, il reclutamento andrebbe condotto ammettendo i più bisognosi.

Madri sole

Fig. 7

ANALISI DI IMPATTO DELLA SUPPORTED WORK DEMONSTRATION (Ore lavorate per mese, campione delle madri non-sposate)

TREND IN HOURS WORKED PER MONTH: AFDC SAMPLE



Note: All experimental-control differentials are significant at the 5 percent level.

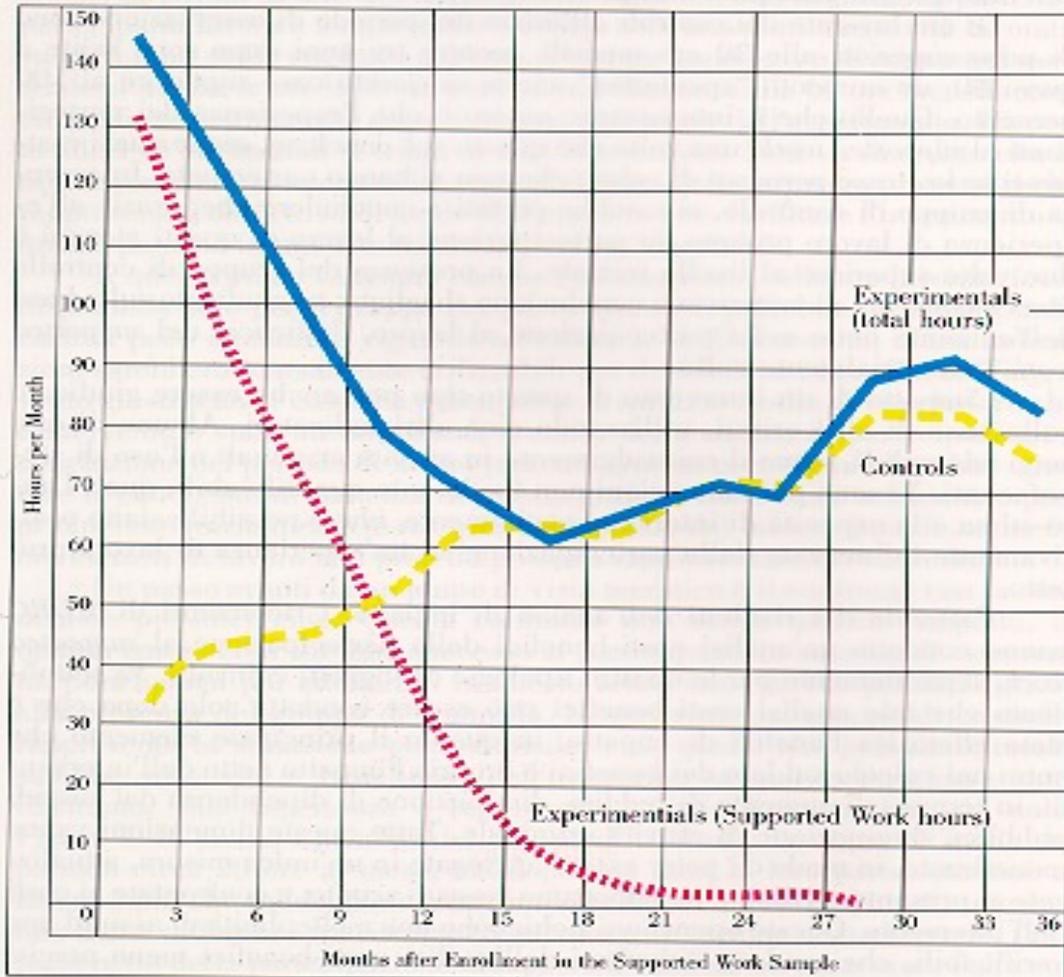
Fonte: Manpower Demonstration Research Corporation (1980)

- ore di lavoro totali medie del gruppo degli esposti
- ore di lavoro "protette" medie del gruppo degli esposti
- ore di lavoro medie del gruppo di controllo

Fig. 8

ANALISI DI IMPATTO DELLA *SUPPORTED WORK DEMONSTRATION*
(Ore lavorate per mese, campione dei giovani *drop-outs*)

TREND IN HOURS WORKED PER MONTH: YOUTH SAMPLE



Note: Experimental-control differentials are significant at the 5 percent level for months 1 to 12.

Giovani drop-outs

- ore di lavoro totali medie del gruppo degli esposti
- ⋯ ore di lavoro "protette" medie del gruppo degli esposti
- - - ore di lavoro medie del gruppo di controllo

NB!!

Cosa mostrano i risultati della SWD?

- Il SW non funziona con i drop-outs
- Piuttosto ragionevole, ma vi sono alcune minacce alla **validità interna** ed **esterna** dei risultati:

Minacce alla validità interna

- **control group substitution bias**: se gli esclusi dal SW reagiscono all'esclusione (\equiv gli assegnati al gruppo di controllo si comportano diversamente da come si sarebbero comportati in assenza della 'demo') cercando di accedere ad iniziative pubbliche/private alternative al SW, con finalità simili, il confronto esposti/non esposti sottostima l'impatto. Viene intaccata la validità interna.

- **attrition bias**: si è in presenza di *attrition* se per qualche motivo una frazione di unità campionarie viene intervistata un numero di volte inferiore alle 5 previste. In generale la presenza di *attrition* (non casuale) compromette i risultati della randomizzazione, e quindi la validità interna (per es: una frazione di unità viene intervistata nelle prime due occasioni e poi viene perduta).

Attrition

Nel caso in questione le ipotesi sul processo generatore dell'*attrition* possono essere molteplici:

1. è plausibile che gli esposti siano più disponibili all'intervista dei non esposti.
 2. è plausibile che tra le cause di mancate reperibilità vi sia la mobilità alla ricerca di un lavoro: i più mobili sono tendenzialmente i più motivati/disponibili al lavoro.
- Per effetto dell'*attrition* la differenza tra esposti e non esposti può essere *sovrastimata*.
 - Se l'indagine è ben progettata e adeguatamente finanziata i danni prodotti da *control group substitution bias* e da *attrition bias* possono essere contenuti:
 - rilevazione al tempo t di ciò che gli intervistati hanno fatto tra $t-1$ e t
 - inseguimento delle unità non reperibili (costoso!)

Minacce alla validità esterna: P' rappresenta P ?

- Tra gli aventi diritto all'ammissione al SW sono presi in considerazione solo coloro che hanno presentato domanda.

A priori, è difficile dire se e in che misura ciò possa costituire un problema rilevante.

Chi sono gli appartenenti a P e a P' ?

Se la selezione che fa passare la popolazione da P a P' è casuale, ciò non costituisce problema; lo è nel caso opposto.

- La popolazione P' "sollecitata" dalla 'demo' rappresenta la popolazione che verrebbe "sollecitata" da un intervento a regime?

Minacce alla validità esterna: P' rappresenta P?

Supponiamo che:

non vi sia *control group substitution bias*

non vi sia *attrition bias*

coloro che presentano la domanda siano rappresentativi di P

I risultati ottenuti in una *demonstration* che coinvolge qualche migliaio di unità consentono di affermare che un intervento a regime sull'intera P otterrebbe gli stessi risultati?

Dipende dalle dimensioni di P in rapporto alle dimensioni del mercato del lavoro che dovrebbe accogliere i beneficiari:

se P è piccolo in rapporto al mercato del lavoro gli effetti di scala sono trascurabili

se P è grande, possono non esserlo.

NB: c'è asimmetria tra 'demo' di successo e 'demo' fallite:
il problema degli effetti di scala si pone se la 'demo' depone a favore della politica in questione.

Valutazione d'impatto con *metodo sperimentale*

- ⇒ impatto medio nella pop.ne stimato dal confronto tra medie variabile risultato trattati e non trattati
- ⇒ problema del *selection bias* risolto (\equiv eliminato) *ex-ante*:
 - *si estrae casualmente un campione di unità dalla popolazione obiettivo;*
 - *si bipartisce casualmente il campione nei due gruppi costituiti da esposti e non esposti.*

Alcuni commenti:

- randomizzazione **non risolve** la non identificabilità a livello del singolo individuo. Y^T continua ad essere osservabile **solo per gli esposti** e Y^{NT} **solo per i non esposti**
- la randomizzazione consente solo di dire che **se** si riscontrano differenze tra il gruppo degli esposti e il gruppo dei non esposti, **allora tali differenze sono dovute alla politica**, vale a dire è identificabile l'impatto medio nella popolazione (funziona con gli impatti medi, grazie alla linearità dell'operatore valor medio)
- **randomizzazione non è sempre praticabile**

Disegno *sperimentale* e situazioni reali

Nella realtà è **difficile ottenere** un disegno sperimentale:

- la randomizzazione può essere non giustificata dal punto di vista etico;
- può essere politicamente difficile fornire un intervento ad un gruppo di soggetti e negarlo ad un altro con gli stessi requisiti di ammissibilità;
- è difficile assicurare che l'assegnazione ai due gruppi sia veramente casuale;
- è possibile che gli esposti escano dal programma e che i non esposti invece vi entrino oppure che individui nel gruppo di controllo partecipino a programmi simili. Tali fenomeni possono invalidare o contaminare i risultati della valutazione;
- nel caso di interventi universali, la natura stessa dell'intervento esclude che vi possa essere un gruppo di controllo;
- i disegni sperimentali possono essere molto costosi sia dal punto di vista economico che dei tempi di realizzazione.

Disegni non *sperimentali*

Se selezione condotta da **un'autorità esterna** o **autoselezione**:

⇒ **disegni non-sperimentali**

il valutatore non può 'manipolare' il processo di generazione dei dati, ma si limita ad osservare quel che accade, dispone cioè di dati **osservazionali** (\equiv *observational data*).

Obiettivo: trovare un'adeguata approssimazione del controfattuale (per confrontarla col fattuale per gli esposti).

Ma ora si deve:

- operare *ex-post*, per l'appunto con dati osservazionali;
- utilizzare in maniera appropriata dati e metodi per (cercare di) "controllare/eliminare" al meglio il *selection bias* e/o la distorsione da *dinamica spontanea*.
- migliorare la **disponibilità di dati** al fine di testare il più possibile le assunzioni a cui si fa riferimento

Metodi non sperimentali

- assegnazione trattati/non trattati non chiara:
con **assunzioni** (non sempre verificabili)
 - *Differenze nelle differenze* (Differences-In-Differences)
 - *Metodi di matching*
- ricorso a qualche fonte di variazione esterna che determina lo status di trattato/non trattato
 - individuare una variabile che influenzi la probabilità di esposizione alla politica e il cui valore sia assegnato in modo assimilabile ad una assegnazione casuale (\equiv in modo indipendente dal processo che genera i risultati potenziali), ovvero esterno all'individuo e non associato alle sue caratteristiche ma che influenza(determina) la sua partecipazione al trattamento
metodo (stimatore) delle variabili strumentali (IV)
- anche altri metodi
 - regression discontinuity design: basato su valori continui di un 'indice di partecipazione' e che le unità siano simili nella vicinanza (sopra e sotto) della soglia definita per partecipare