
Bias e Confondimento: perché gli uomini
fanno piú incidenti delle donne?

Alessio Farcomeni
Sapienza - Università di Roma

Associazione e studi campionari

- Facciamo studi campionari per fare inferenza su *relazioni* (es. tra esposizioni ed eventi).
- Lo zero non esiste: qualunque analisi mostrerá sempre un effetto non nullo sul campione. Il problema é trovare una spiegazione della relazione trovata.
- Quattro alternative possibili:
 1. Caso
 2. Bias
 3. Confondimento
 4. Relazione effettiva (sulla popolazione)

Come dimostrare una relazione effettiva

- Escludendo tutte le altre spiegazioni possibili: caso, bias, confondimento.
- Due parole sul caso (che non é argomento di oggi)
- Es: i laureati dell'acquario negli USA sono l'8%. Quelli dei gemelli solo il 6%.
- Direi che non possiamo escludere il caso (anzi, é tutto lí)

Come escludere il caso (formalmente)

- Si specifica una ipotesi nulla H_0 di assenza di effetto.
- Si calcola il p -value: la probabilità che l'effetto osservato o uno più estremo sia ottenuto per effetto del caso solamente (ovvero, condizionatamente ad H_0).
- Se p é piccolo (minore di un α pre-specificato, es. $\alpha = 5\%$) si può escludere il caso. Altrimenti *non* si può escludere il caso (non abbiamo comunque dimostrato equivalenza, ma questa é un'altra storia)
- Su un campione di $n = 1500$ americani, il test del chi-quadrato per la relazione tra segno zodiacale e titolo di studio da $p = 0.604$.

Menú

- Cosa é il bias
- Come escludere il bias in uno studio
- Cosa é il confondimento
- Come escludere il confondimento in uno studio

Cosa é il bias

- Qualunque fonte di errore nella determinazione dell'associazione di interesse.
- Si possono avere errori da come i partecipanti sono inseriti nello studio (bias di selezione)
- Si possono avere errori da come l'informazione é misurata sui partecipanti (bias di osservazione)

Escludere il bias

- Esaminare uno per volta una lista di bias possibili (prossimi lucidi)
- Molti tipi non sono curabili una volta presenti: prevenire in fase di disegno dell'esperimento!
- L'esclusione del bias a posteriori viene spesso semplicisticamente fatta tramite mere riflessioni su ciascun tipo di bias, possibilmente corredate da analisi ad-hoc.
- In caso il bias non possa essere escluso, é talvolta possibile usare modelli statistici avanzati. Altrimenti, si buttano i dati (o ci si fa dell'altro).

Esempio

- Esempio: bias da dati mancanti. Quanto guadagni?
- Per argomentare a sfavore di questo bias, mostro che la percentuale di dati mancanti non é significativamente differente tra esposti/non-esposti e tra casi/controlli. Ad esempio, l'indicatore di dati mancanti non é associato al titolo di studio, ai mQ di casa dove si abita.
- Se non sono pochi e si ha il sospetto siano informativi, esistono modelli formali che trattano i dati mancanti. Parlatene con uno statistico medico.

Esempio di bias da dati mancanti

- Valutazione degli aspetti multidimensionali del dolore su 108 primipare, tramite somministrazione del questionario di Gill a inizio (4cm) e fine (8cm) travaglio.
- Un test sull'intensità del dolore dimostrava che a fine travaglio si soffre *significativamente* di meno che a inizio travaglio.
- Questionario di Gill: 72 domande.
- Ripetizione dello studio con un questionario ridotto (12 domande) mostrava esattamente l'effetto opposto.

Selection bias

- Selection bias: il campione é rilevato da una sottopopolazione della popolazione di interesse.
- Ovvero: l'inclusione dei soggetti é dipendente dalle loro caratteristiche (esposizione, evento, altro).
- Es: studio sulle tossicodipendenze che include solo soggetti nei SeRT.

Particolari tipi di bias di selezione

- Bias da reclutamento dei casi. Es. reclutiamo piú facilmente soggetti con forme piú gravi della patologia.
- Bias del volontario: i controlli sono reclutati in una popolazione non rappresentativa della popolazione sana. Es: soggetti afferenti a strutture sanitarie
- Diagnostic bias: la diagnosi é differenziale tra esposti e non. Es: relazione tra contraccettivi orali e tumore endometriale. I contraccettivi orali danno piú frequentemente emorragie uterine, la cui investigazione porta a facilitare la diagnosi di neoplasia.
- Attrition bias: perdite al follow-up informative.

Esempio di attrition bias

- Conteggio CD4 in pazienti HIV a 6, 12, 24 e 36 mesi dal primo trattamento.
- I pazienti possono uscire dallo studio per morte.
- I pazienti con CD4 basali piú basse hanno maggiore rischio di uscire dallo studio.
- Saranno quindi sempre meno rappresentati.
- Soluzione complessa: Joint-Models.

Bias di osservazione

- Errori di misura differenziali (soluzione: standardizzare la raccolta dati!)
- Recall bias: differente “ricordo” degli eventi tra i gruppi
- Misclassificazione di casi o controlli. Se non differenziale, fa sottostimare la differenza tra i gruppi. Se differenziale, può far sotto o sovrastimare.
- Bias da dati mancanti: se la mancanza del dato dipende dal valore che avrei misurato o da predittori non misurati.
- Interviewer bias: l’intervistatore proietta le proprie aspettative sull’intervistato.

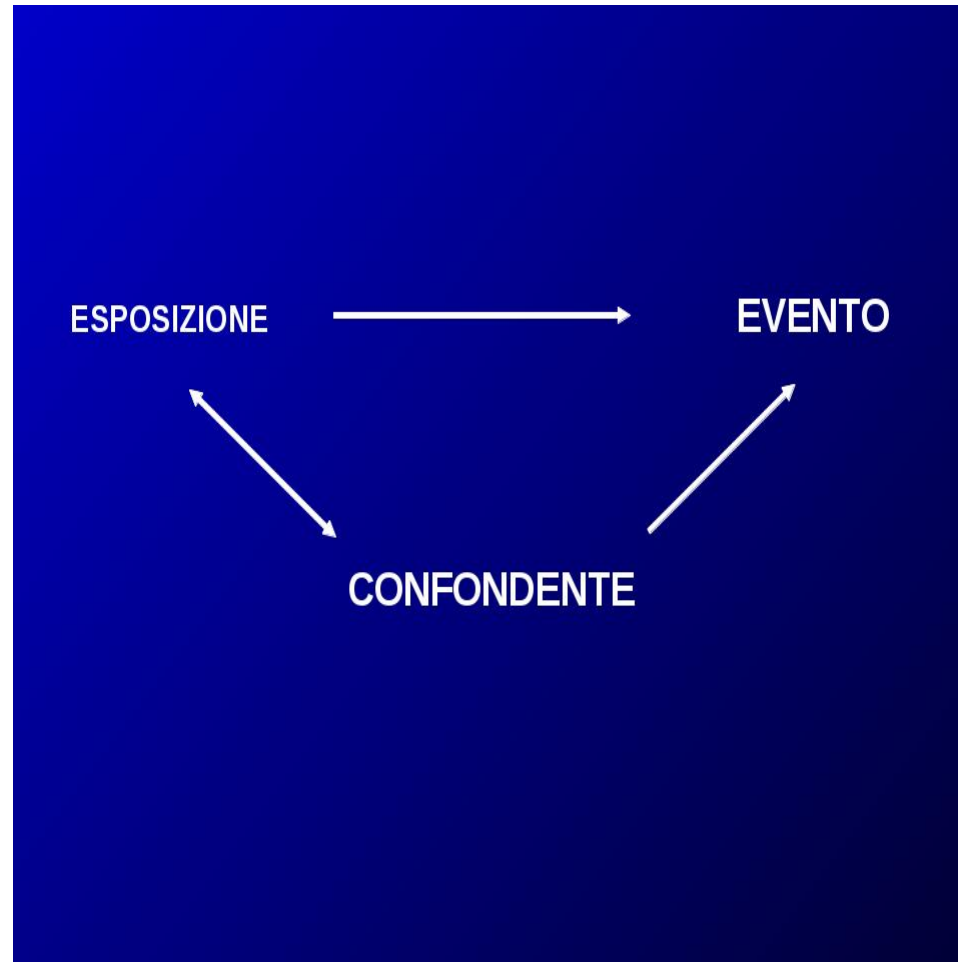
Recall bias

- Un esempio di recall bias é quello relativo a studi sulla demenza.
- Soggetti che hanno sperimentato l'evento possono ricordare meno bene l'esposizione a certi fattori di rischio.

Cosa é il confondimento

- Una mistura di effetti rispetto a una terza varaibile (o piú)
- Il confondente deve essere associato *sia* con l'esposizione che con l'evento.
- Il confondente é responsabile in parte o *totalmente* dell'associazione vista nei dati!

Graficamente

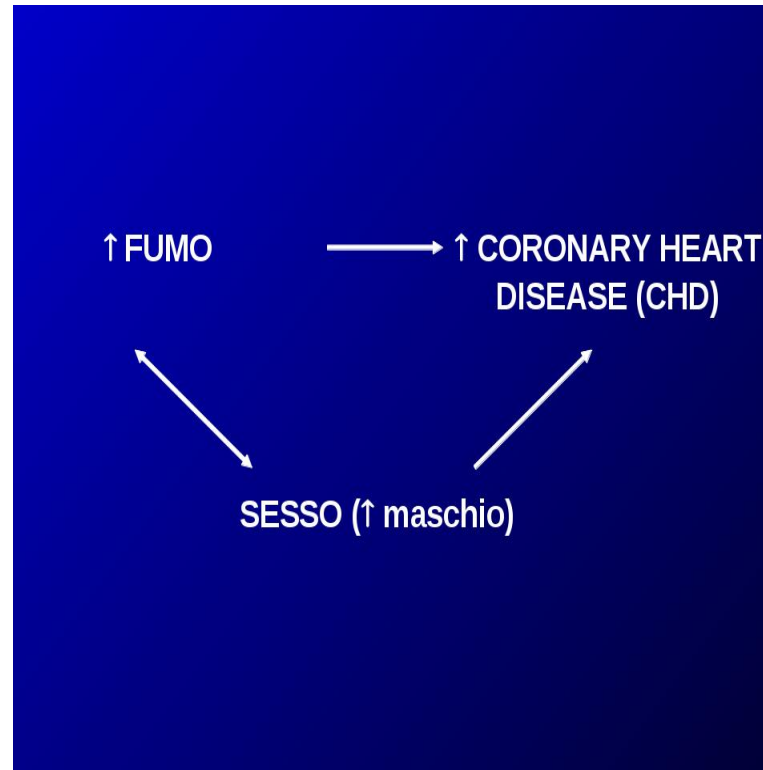


Piú esplicitamente

Un confondente puó

- mascherare una associazione esistente
- mostrare una associazione inesistente
- invertire la direzione di una associazione
- far sovrastimare la forza di una associazione
- far sottostimare la forza di una associazione

Esempio grafico



Il confondente porta a sovrastimare l'effetto del fumo.

Cosa é non é un confondente

- Se una variabile é associata con l'esposizione ma non con l'evento, non é un confondente.
- Es. relazione tra fumo e tumore ai polmoni. L'alcol é associato al fumo ma non é un confondente.
- Se una variabile é associata con l'evento ma non con l'esposizione, non é un confondente.
- Fattori intermedi (nella catena causale)
- Es. relazione tra Vit. D e 1,25-DHCC. 25-HCC non é un confondente.

Escludere il confondimento

- Il confondimento é curabile a patto che si sia misurato il confondente.
- Se non si ha una lista affidabile di confondenti, misurare quantomeno tutti i fattori di rischio noti. Alcuni saranno confondenti, altri no.
- La lista di confondenti non é mai esaustiva. É anche per questo che gli studi osservazionali non dimostrano la causalitá.

Escludere il confondimento

- Negare: dimostrare che i possibili confondenti misurati non sono associati con l'esposizione o con l'evento.
- Se presente:
 - ▷ Restrizione, stratificazione
 - ▷ Appaiamento (o randomizzazione se possibile)
 - ▷ Analisi multivariata (aggiustare la relazione, effetto "al netto dei confondenti")
- Sempre: argomentare a sfavore dell'effetto di confondenti non noti. Unica soluzione per quelli: randomizzazione.

Esempio: relazione tra pillola e infarto

- RR: 2.0
- RR nelle non fumatrici: 1.6 RR nelle fumatrici: 18.0
- RR nelle bevitrici: 2.2 RR nelle non bevitrici: 1.9
- Stratificare rispetto al fumo, l'alcol non é probabilmente un confondente (testare questa ipotesi).
- RR aggiustato per età: 2.5

Quando stratificare e quando aggiustare

- Se l'effetto del confondente é forte, stratifico (es. associazione tra pillola e infarto nelle donne sopra i 35 anni. Confondente: fumo)
- Se l'effetto del confondente é debole, aggiusto (es. associazione tra fumo e infarto per età nella fascia 30-50 anni)

L'esempio del titolo...

- Qualunque studio osservazionale mostrerà sempre che gli uomini fanno piú incidenti delle donne.
- Tranne uno: quello che studia la relazione tra sesso e rischio di incidente.
- Perché?

Confondimento

- Per ragioni socioeconomiche, gli uomini passano piú tempo in macchina per le donne.
- Se appaiamo per tempo speso in macchina, non troviamo alcun effetto: gli uomini e le donne hanno approssimativamente lo stesso rischio di incidente.
- Le assicurazioni fanno comunque bene a far pagare piú gli uomini delle donne. Perché?

Associazione vs Causalità

- Alle assicurazioni non interessa perché questo succede, e rilevare il sesso é molto piú sicuro che rilevare il tempo speso in auto.
- In ambito sanitario e clinico, a noi interessa la causalità: cosa succede se intervengo. Se cambiamo sesso a un individuo ma il suo tempo speso in macchina resta uguale, il rischio di incidente rimane invariato.

I trial clinici

- La randomizzazione garantisce l'assenza di confondimento.
- L'assenza di compliance può portare a confondimento!
- Validità interna: assenza di bias di selezione (nel reclutamento o nell'assegnazione), assenza di bias di performance (cura differenziale) e altri bias
- Validità esterna: possibilità di generalizzare (criteri di inclusione ed esclusione, regime di trattamento, outcome)

Sintesi del Menú

- Cosa é il bias: ho misurato male o selezionato male il campione.
- Come escludere il bias in uno studio: prevenire.
- Cosa é il confondimento: fattori esterni influenzano la relazione di interesse.
- Come escludere il confondimento in uno studio: restrizione, stratificazione, matching, analisi multivariata.

Sintesi e precisazione

- Abbiamo detto che qualunque studio statistico mostrerà sempre una relazione non nulla. Magari piccola, ma non nulla.
- Quattro alternative possibili: caso, bias, confondimento, relazione effettiva (sulla popolazione).
- Precisazione: gli effetti non sono *alternativi*, ma sono tutti presenti, si sommano e interagiscono.
- Escludere caso, bias e confondimento significa mostrare che questi tre fattori hanno un effetto minimale rispetto all'effetto della relazione effettiva.

Conclusione

... il che é un altro modo per concludere come abbiamo cominciato: lo zero non esiste.