

LEZIONI DI EPIDEMIOLOGIA

Dott. Simone Accordini

Lezione n. 5:

- Il confondimento
- I bias



Sezione di Epidemiologia & Statistica Medica
Università degli Studi di Verona

L'INTERPRETAZIONE CAUSALE DI UN'ASSOCIAZIONE EMPIRICA

REALTA' (ignota)

$$\begin{array}{l} D_1 \Rightarrow \pi_1 \\ D_0 \Rightarrow \pi_0 \end{array} \Rightarrow \frac{\pi_1}{\pi_0} = \Psi$$

vero valore
del parametro di effetto

RICERCA EMPIRICA

$$\begin{array}{l} D_1 \Rightarrow P_1 \\ D_0 \Rightarrow P_0 \end{array} \Rightarrow \frac{P_1}{P_0} = \hat{\Psi} = RR$$

processo
di stima

Possibili errori nel processo di stima:

Tipo I: osserviamo un'associazione quando in realtà non esiste

$RR > 1$ o $RR < 1$ ma $\Psi = 1$

Tipo II: non osserviamo alcuna associazione quando in realtà esiste

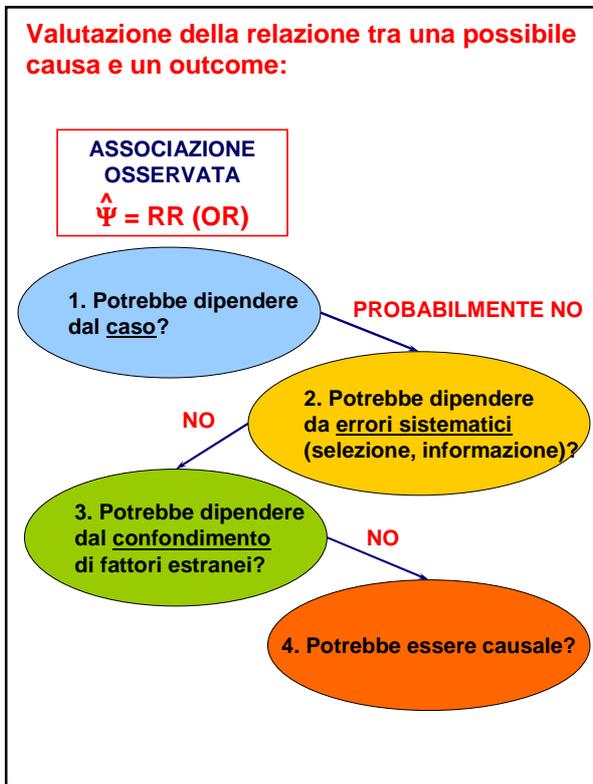
$RR = 1$ ma $\Psi > 1$ o $\Psi < 1$

FONTI di ERRORE
(possono causare errori di tipo I e II)

1. VARIABILITÀ CASUALE
scarto tra la stima (RR) e il vero valore del parametro (Ψ) dovuto al caso
→ dipende dalla numerosità del campione (decrece al crescere del numero di soggetti arruolati nello studio)
→ deve essere misurata

2. VARIABILITÀ SISTEMATICA (BIAS)
deviazione sistematica dovuta ad errori nella selezione dei soggetti e/o nella raccolta dell'informazione
→ dipende dal disegno e/o dalla conduzione dello studio
→ deve essere prevenuta

3. CONFONDIMENTO
deviazione sistematica dovuta alla differente distribuzione di fattori associati all'occorrenza della malattia nel gruppo esposto e non esposto
→ dipende dalle relazioni complesse tra le diverse esposizioni e la malattia
→ deve essere controllato



CONFONDIMENTO

gruppo non esposto \neq gruppo esposto

sbilanciamento nella distribuzione
di fattori associati all'occorrenza
della malattia (confondenti)

\Rightarrow la differenza nella frequenza dell'outcome è prodotta
da un mix di effetti diversi (esposizione e confondenti)

relazione condizionale al confondente (C):

relazione valutata entro gruppi omogenei rispetto a C
 \Rightarrow priva di confondimento

\Downarrow

una relazione d'occorrenza può essere interpretata
in senso causale se è condizionata al set dei possibili
confondenti:

$$P = f(D|C)$$

$C = (C_1, C_2, \dots, C_k)$ = insieme dei confondenti

Esempio:

Confronto tra l'incidenza di tumore polmonare in un gruppo di operai maschi (40-50 anni) in una fonderia e l'incidenza di tumore polmonare in un gruppo di impiegati dello stesso sesso ed età.

	casi	persone-anno
operai	65	10.000
impiegati	35	10.000

RR = 1.86 (relazione grezza)

... ma l'80% degli operai fumano contro il 20% degli impiegati.
I dati disaggregati per abitudine al fumo sono:

FUMATORI

	casi	persone-anno
operai	60	8.000
impiegati	15	2.000

RR = 1

NON FUMATORI

	casi	persone-anno
operai	5	2.000
impiegati	20	8.000

RR = 1

(relazioni condizionali alle categorie del fumo)

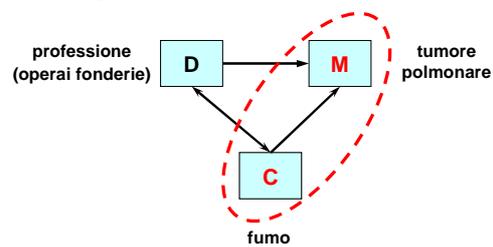
CARATTERISTICHE DI UN CONFONDENTE:

1. non deve dipendere dall'esposizione o dalla malattia



2. deve essere un fattore di rischio della malattia

→ deve essere associato alla patologia nel gruppo non esposto

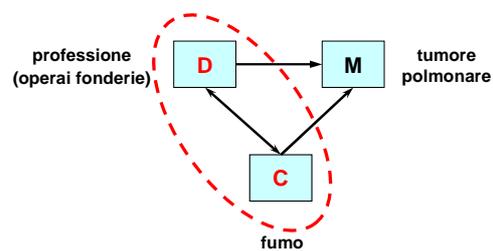


Il fumo ha un effetto sull'incidenza del tumore polmonare tra i non esposti (impiegati): determina un aumento dell'incidenza della malattia.

3. deve essere associato all'esposizione nella popolazione sorgente dei casi

→ in entrambe le coorti (esposta e non esposta) all'inizio del follow-up [studio di coorte]

→ nei controlli come surrogato della popolazione sorgente [studio caso-controllo]



Il fumo è associato con la professione: gli operai della fonderia includono una proporzione maggiore di fumatori (80%) rispetto agli impiegati (20%).

⇒ Gli operai della fonderia presentano un'incidenza di tumore polmonare maggiore degli impiegati per effetto del fumo.

Esempio: Studio caso-controllo sulla relazione tra alto consumo di zucchero (C.Z.) e morbo di Crohn.

relazione grezza tra D e P

	Casi	Controlli
Alto C.Z.	257	200
Basso C.Z.	243	300
	500	500

OR = 1.59

relazioni condizionali tra D e P

Fumatori

	Casi	Controlli
Alto C.Z.	200	100
Basso C.Z.	100	50
	300	150

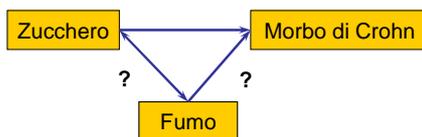
OR = 1.00

Non fumatori

	Casi	Controlli
Alto C.Z.	57	100
Basso C.Z.	143	250
	200	350

OR = 1.00

Il fumo è un confondente della relazione tra l'alto consumo di zucchero e l'occorrenza del morbo di Crohn?



Esempio (continua):

	Fumatori	Non fumatori
Alto C.Z.	100	100
Basso C.Z.	50	250
	150	350

OR = 5.00

associazione tra alto consumo di zucchero e fumo nella popolazione sorgente (controlli)

	Casi	Controlli
Alto C.Z.	200	100
Basso C.Z.	100	50

associazione tra fumo e malattia nei non esposti

	Casi	Controlli
Alto C.Z.	57	100
Basso C.Z.	143	250

	Casi	Controlli
Fumatori	100	50
Non fumatori	143	250
	243	300

OR = 3.50

```

    graph TD
      Zucchero --> Morbo di Crohn
      Fumo --> Morbo di Crohn
      Zucchero --> Fumo
      Fumo --> Zucchero
  
```

OR = 5.00 OR = 3.50

⇒ il fumo è un confondente

METODI PER CONTROLLARE IL CONFONDIMENTO:

Nel disegno dello studio

- **restrizione:**

uso di criteri di ammissibilità nella selezione dei soggetti
(*restrizione ad un range di età, un sesso, ...*)

- **matching:**

selezione del gruppo di confronto (non esposti - studio di coorte, controlli - studio caso-controllo) in modo che sia 'uguale' al gruppo di interesse (esposti, casi) rispetto alla distribuzione dei potenziali confondenti

- **randomizzazione (solo nelle sperimentazioni):**

assegnazione casuale dei soggetti al gruppo sperimentale e di controllo

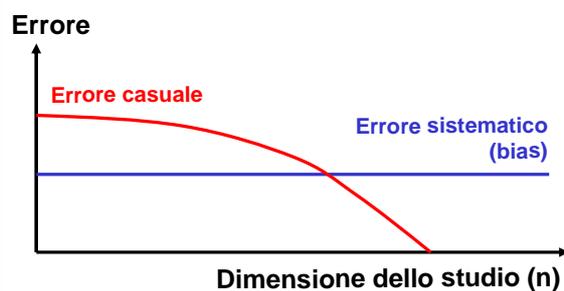
Nell'analisi dello studio

- **stratificazione**

- **modelli multivariati**

VARIABILITA' SISTEMATICA (BIAS)

qualsiasi errore sistematico (non dovuto al caso) che risulta in una stima non corretta dell'associazione tra esposizione e rischio della malattia



ERRORI SISTEMATICI (⇒ errori di tipo I e II):

- nella selezione della popolazione in studio (**selection bias**)
- nella misurazione dello status di esposizione o di malattia (**information bias**)

SELECTION BIAS

Esempio (studio caso-controllo):

Relazione tra elevato consumo di alcool e insorgenza di cirrosi epatica (studio 'hospital-based').

Quanto i pazienti ospedalizzati per trauma (controlli A) sono rappresentativi della popolazione sorgente dei casi?

	Casi (cirrosi)	Controlli A (trauma)	Controlli B (non-trauma)
Elevato consumo di alcool	80	40	10
Modesto / nessun consumo di alcool	20	60	90

OR = 6.00

OR = 36.00

Nei controlli A la frequenza dell'esposizione è maggiore che nella popolazione sorgente dei casi

selection bias ⇒ sottostima dell'associazione

Esempio (studio di coorte):

Relazione tra fumo e insorgenza di tumore al polmone.

	Tumore al polmone		
	sì	no	
Fumatori	9	91	100
Non-fumatori	1	99	100

RR = 9/100 : 1/100 = 9

Si supponga che 33 fumatori sportivi non vogliono partecipare allo studio:

	Tumore al polmone		
	sì	no	
Fumatori sportivi	0	7	7
Fumatori non sportivi	9	51	60
Non-fumatori	1	99	100

selection bias ⇒ RR = 9/67 : 1/100 = 13.4

INFORMATION BIAS (MISCLASSIFICAZIONE)

errore di misurazione che consiste nell'osservare il livello di esposizione e/o la categoria dell'outcome erranei

DIFFERENZIALE

- ▷ Dovuta ad errori sistematici
- ▷ Dipende dall'esposizione o dall'outcome
- ▷ E' un bias
- ▷ La misura di associazione è distorta in qualsiasi direzione

NON-DIFFERENZIALE

- ▷ Dovuta ad errori casuali
- ▷ Non dipende dall'esposizione o dall'outcome
- ▷ Non è un bias
- ▷ Usualmente 'spinge' la misura di associazione verso l'effetto nullo (RR=1)

Esempio (studio di caso-controllo) - misclassificazione differenziale:

Relazione tra uso di tabacco, alcool e farmaci e occorrenza di malformazioni nel neonato.

	Madri di	
	bambini con malformazioni	controlli
Uso di tabacco, alcool, farmaci	a	b
Non uso	c	d

• le madri dei bambini con malformazioni (casi) ricordano le passate esposizioni meglio delle madri dei bambini sani (controlli)

• sottostima di 'b' ⇒ **sovrastima dell'associazione (bias)**

Esempio (studio di coorte) - misclassificazione non differenziale:

Relazione tra alcool e insorgenza di tumore alla laringe.

No misclassificazione	Incidenza per milione	RR
1,000,000 bevitori	50	5.0
500,000 nonbevitori	10	

50% dei bevitori misclassificati

500,000 bevitori	50	1.7
1,000,000 "non-bevitori"	30	

media ponderata = $[500,000 \times 10 + 500,000 \times 50] / 1,000,000$

50% dei bevitori e dei non bevitori misclassificati

750,000 "bevitori"	36.7	1.0
750,000 "non-bevitori"	36.7	

$[500,000 \times 50 + 250,000 \times 10] / 750,000$

$[250,000 \times 10 + 500,000 \times 50] / 750,000$