

# Relazione causale

Associazione statistica



Relazione causale

# Problemi di interpretazione della relazione causale

- Due eventi possono essere correlati fra loro in due modi
  - Associazione statistica
  - Relazione causale
- Per esempio se uno studio epidemiologico mostra che una malattia è più frequente tra coloro che possiedono una certa caratteristica si può affermare che esiste una associazione statistica che può essere considerata una presunta relazione causale
- La relazione causale è ciò che accade nel mondo reale
- Una associazione statistica potrebbe anche non essere relazione causale

# Relazione tra risultati di uno studio e realtà

		<b>realtà</b>	
		Relazione causale SI'	Relazione causale NO
<b>studio</b>	Associazione statistica SI'	<b>a</b>	<b>Errore <math>\alpha</math></b>
	Associazione statistica NO	<b>Errore <math>\beta</math></b>	<b>d</b>

# Errori

- Errore casuale dovuto al caso
  - Variabilità casuale
  - Potenza dello studio
- Errore sistematico dovuto ad una non corretta pianificazione/esecuzione/analisi dello studio
  - Bias di selezione
  - Bias di informazione
  - Bias di confondimento

# PRECISIONE

- *Definizione*: correttezza nella misurazione o nella stima di un evento in funzione dell'errore casuale
- In uno studio epidemiologico la variazione casuale ha diverse componenti: le principali sono costituite dal processo di selezione del campione e dalla misura delle variabili

# PRECISIONE

- La precisione può essere incrementata:
  - aumentando le dimensioni del campione; le dimensioni del campione possono essere calcolate con apposite formule in relazione al grado di “*precisione*” desiderato.
  - aumentando l’*efficienza* di raccolta delle informazioni

# INFERENZA CAUSALE

- **Associazione spuria**
- **Associazione indiretta**
- **Associazione causale**

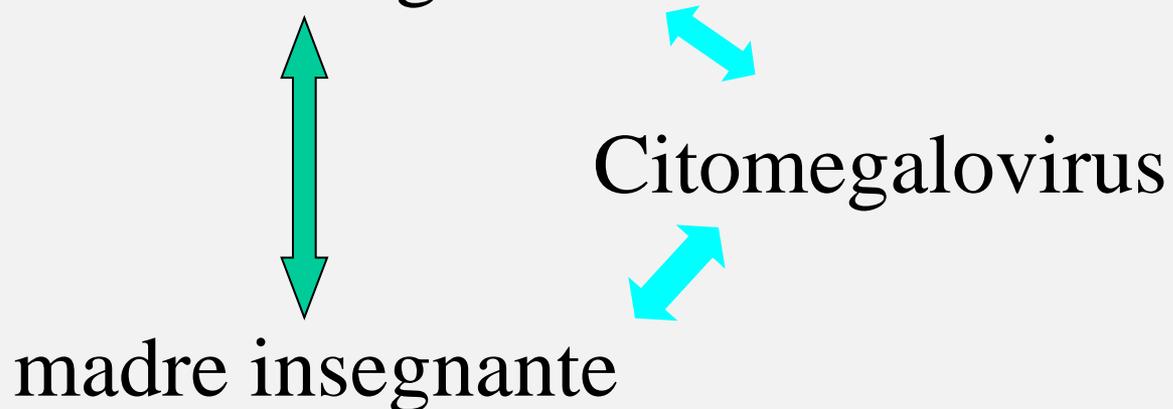
# ASSOCIAZIONE SPURIA

- Relazione fittizia tra le variabili considerate determinata da circostanze aleatorie o da distorsioni metodologiche dello studio
  - Errore di selezione
  - Errore nella raccolta dati

# ASSOCIAZIONE INDIRETTA

- Si ha quando un presunto fattore e una malattia risultano apparentemente in relazione tra loro per l'esistenza di una condizione comune ad entrambi.

Malformazioni congenite



# Associazione causale

- Causa necessaria e sufficiente  
rapporto biunivoco tra causa ed effetto: ogni volta che interviene la causa si verifica l'effetto e viceversa ogni volta che si osserva un effetto deve essere precedentemente accaduta la causa

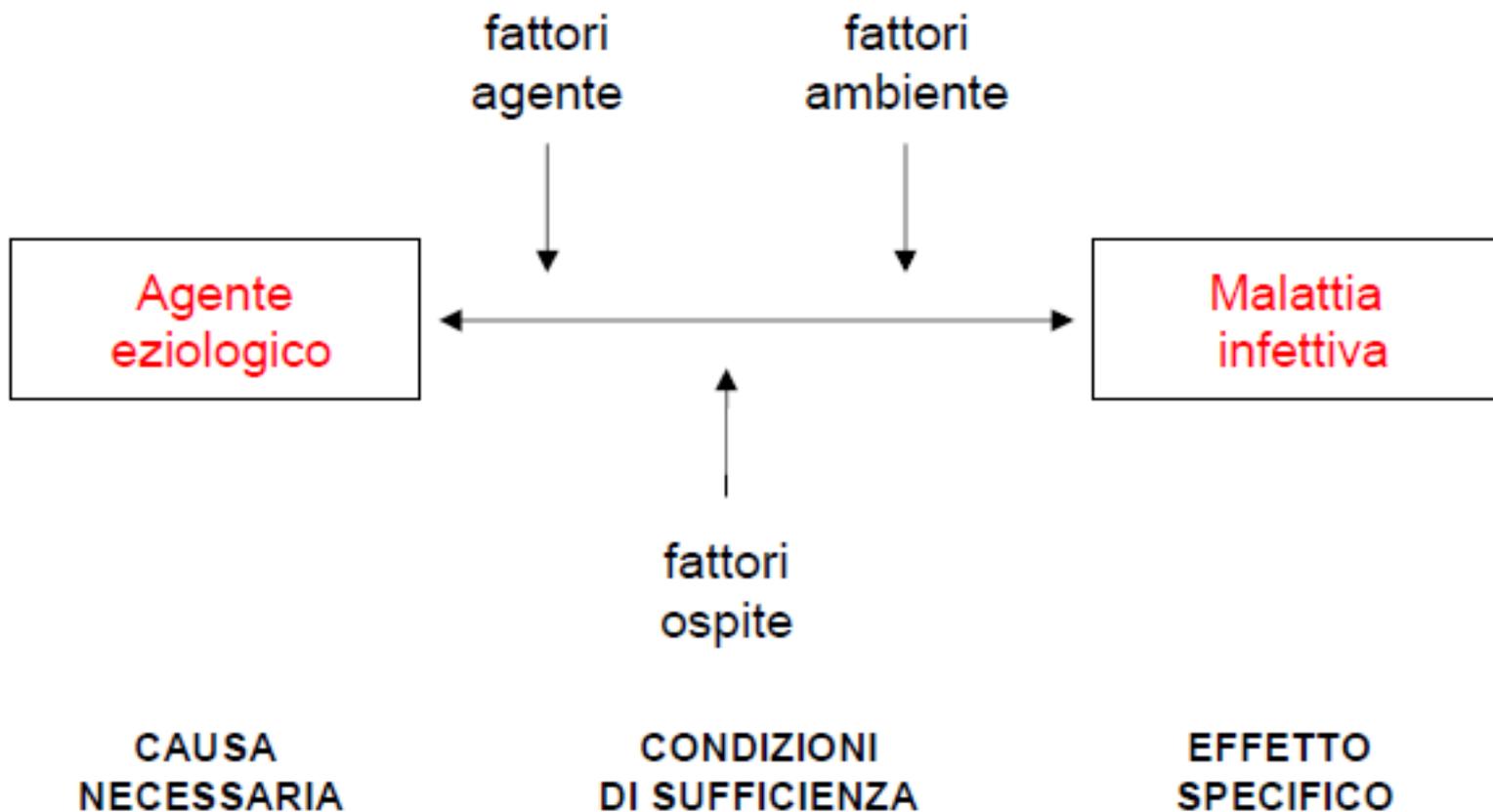
		Malattia	
		presente	assente
Agente causale	presente	a	-
	assente	-	d

# Associazione causale

- Causa necessaria ma non sufficiente  
L'agente etiologico è necessario ma non sufficiente a determinare malattia; esempio concetto di patogenicità delle malattie infettive

		Malattia	
		presente	assente
Agente causale	presente	a	b
	assente	-	d

# Associazione causale

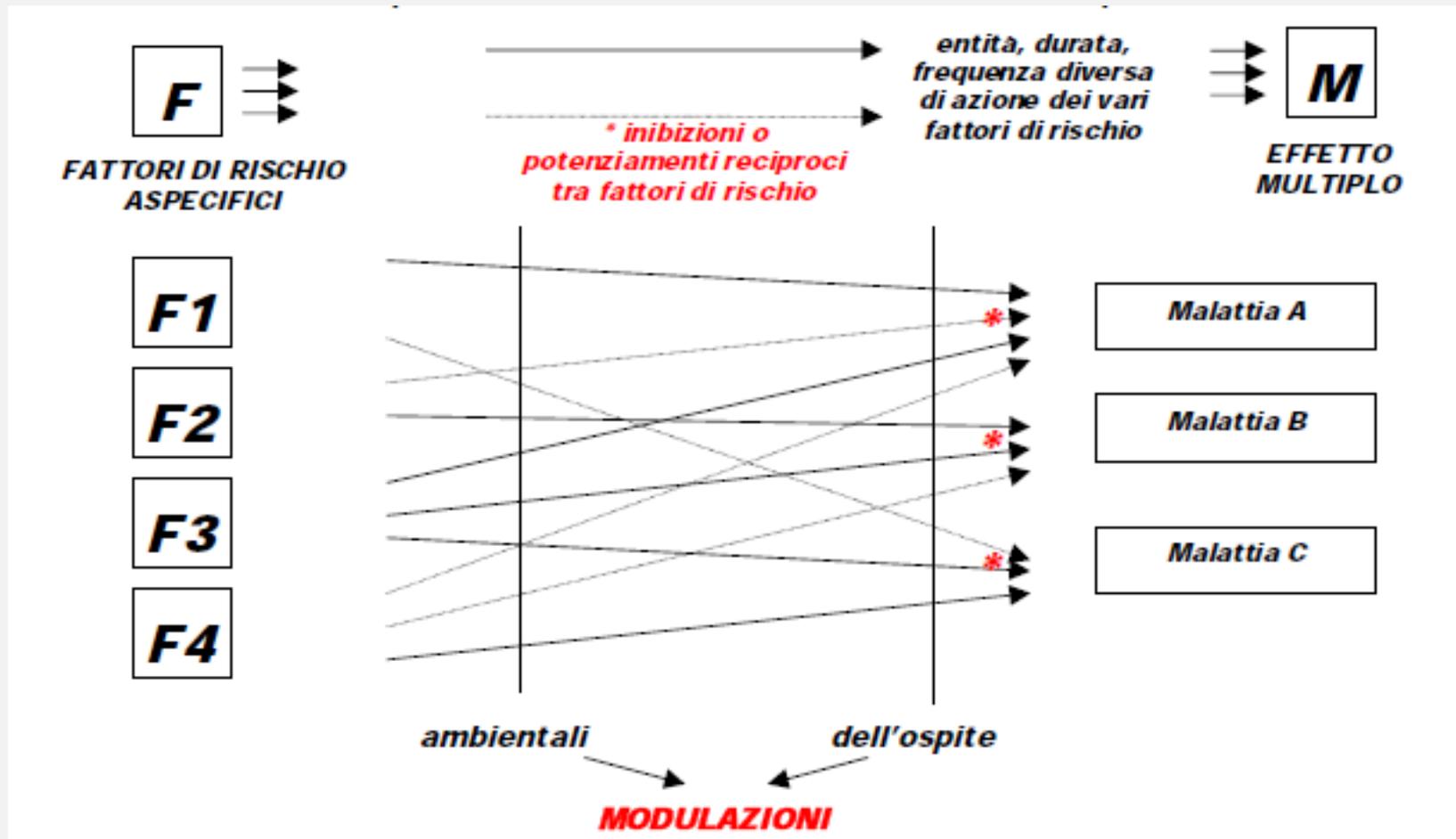


# Associazione causale

- Causa non necessaria e non sufficiente  
l'agente è solo uno dei fattori causali e la sua presenza non è necessaria allo sviluppo della malattia; esempio malattie cronico degenerative e fumo.

		Malattia	
		presente	assente
Agente causale	presente	a	b
	assente	c	d

# Rapporto di rischio aspecifico e multiplo (malattie multifattoriali)

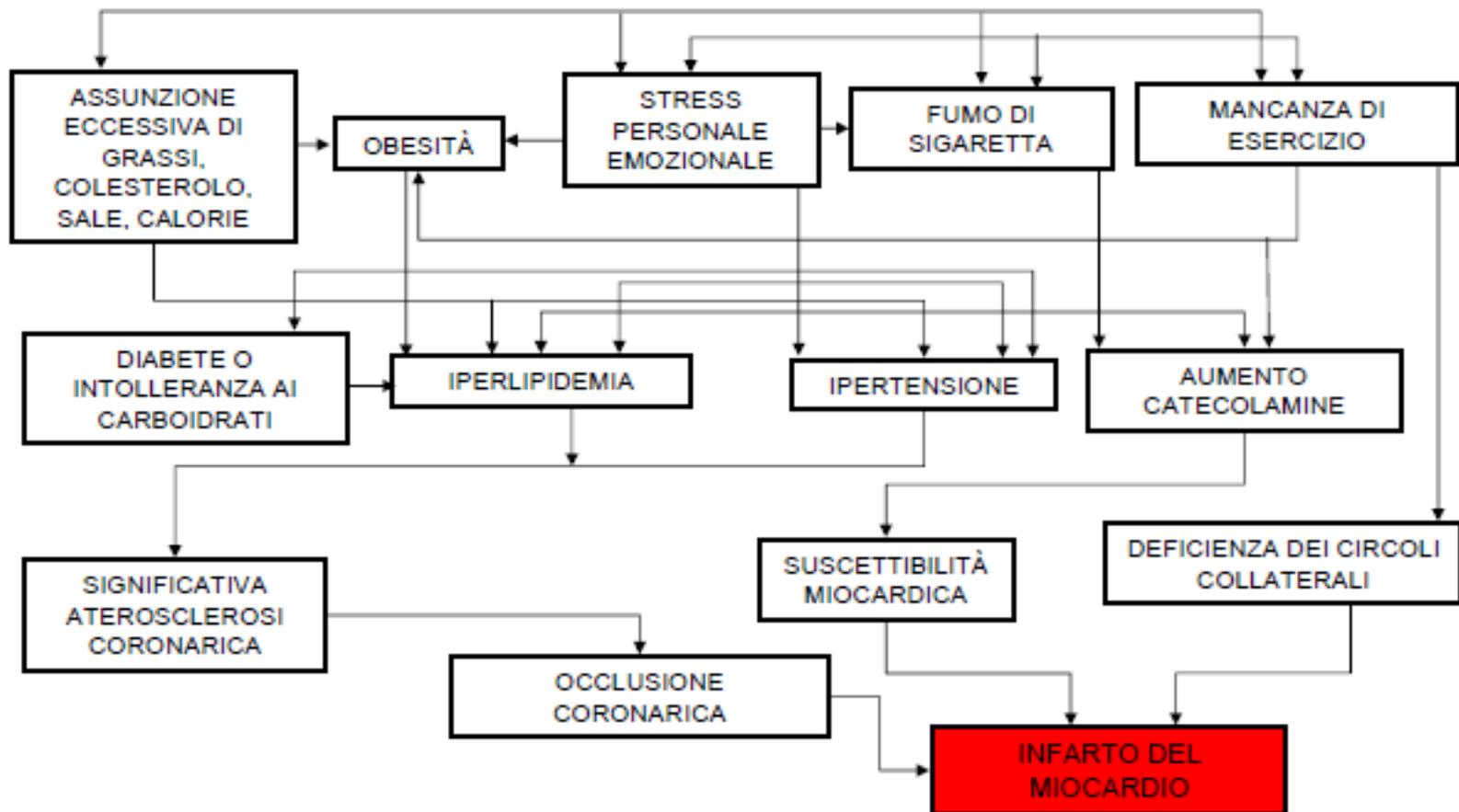


# Causa non necessaria e non sufficiente

- La maggior parte delle situazioni sanitarie di relazione causa effetto
- Il fumo
  - non è sempre in grado di provocare la neoplasia (non sufficiente)
  - si hanno casi di neoplasia in soggetti non fumatori (non necessaria)
- Rete di causazione

# Rete causale

## Rete di causazione per l'infarto miocardico acuto



# Multi-fattorialità: assenza di interazione



# Mortalità (x100.000) e RR per Ca al polmone da fumo e asbesto

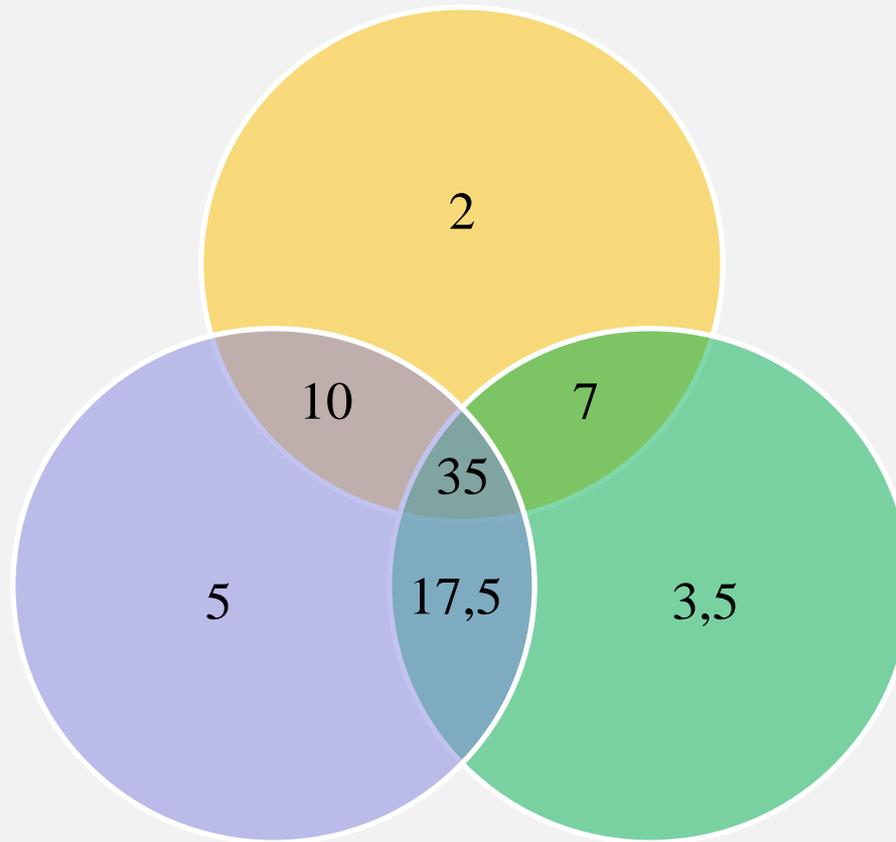
## Tassi di mortalità/100.000

		Fumo	
		No	Si
Asbesto	No	11	123
	Si	58	<b>602</b>

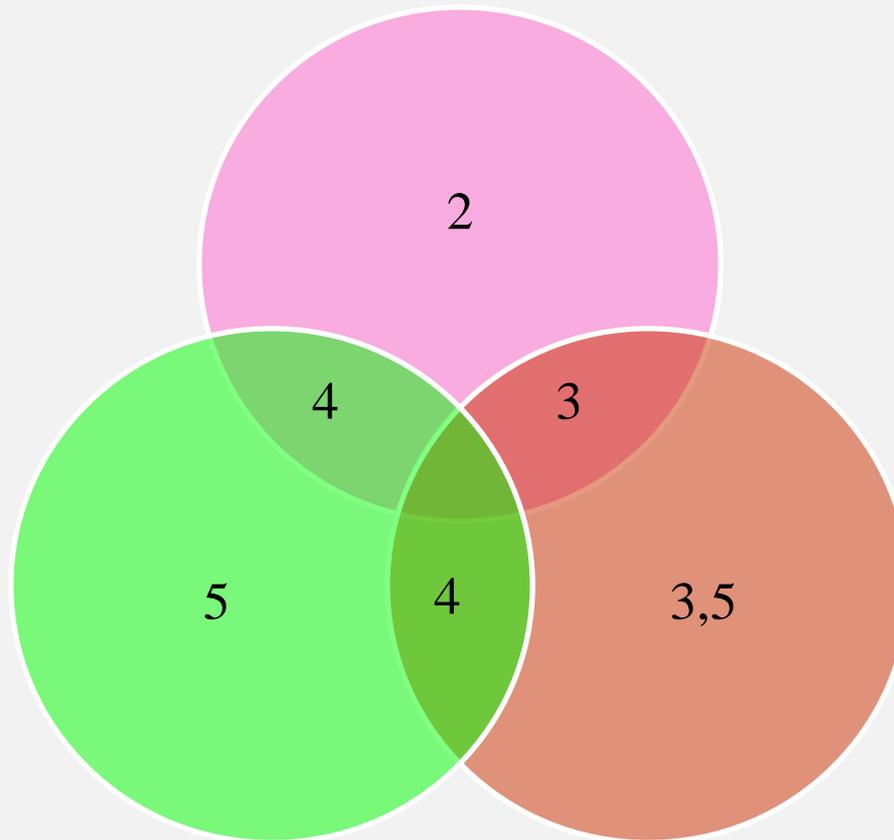
## Rischio relativo

		Fumo	
		No	Si
Asbesto	No	1	11,2
	Si	5,3	<b>54,7</b>

# Multi-fattorialità: Sinergismo



# Multi-fattorialità: Antagonismo



# ASSOCIAZIONE CAUSALE

Criteria per la valutazione di un'associazione

- Forza
- Consistenza
- Specificità
- Congruenza temporale
- Plausibilità biologica

# Forza dell'associazione

- Frequenza di comparsa della malattia nei soggetti esposti al fattore di rischio rispetto a quelli non esposti:
  - Rischio relativo
  - Odds ratio
  - Evidenza statistica
- Relazione dose-risposta

# Consistenza

- Riproducibilità del risultato in diverse popolazioni o con differenti metodologie investigative

# Specificità

- Criterio minore – predittività del determinante nei confronti dell'effetto

# Congruenza temporale

- Il rapporto causale implica l'antecedenza di uno o più fattori rispetto all'effetto.

# Plausibilità biologica

- Meno rigoroso, dipende dalle conoscenze del momento.

# Associazione tra (HPV) DNA e cancro cervicale

	No. studies	HPV DNA prevalence (%)		OR (95% CI)
		Controls	Cases	
Squamous	9	13.4	90.7	83.3 (54.9 to 105.3)
Adeno and mixed	6	15.4	91.9	68.7 (36.2 to 130.5)

CI, confidence interval; IARC, Agency for Research on Cancer OR, odds ratio.

IARC multicentre case-control study

# Rischio relativo e frazione attribuibile di alcuni fattori di rischio

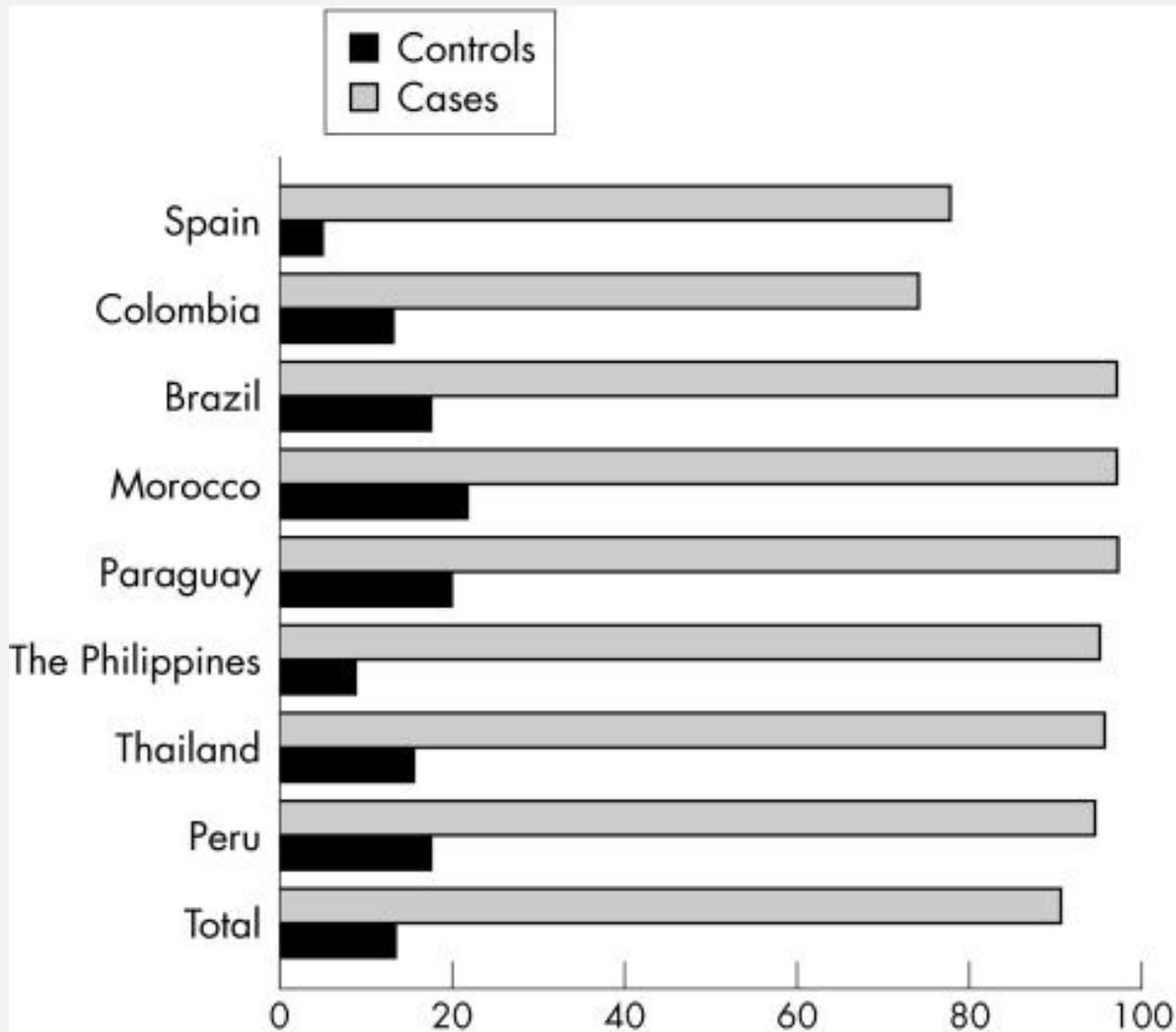
RR/OR		AF%
>500	HPV-DNA-18 and cervical adenocarcinoma in the Philippines	99
	HR-HPV-DNA and cervical cancer in Costa Rica	80
100	HPV-DNA and cervical cancer in Bangkok	90
50	HBsAg and liver cancer in Taiwan	60
20	HBsAg and liver cancer in Greece	60
10	HCV and liver cancer in Italy	40
	Cigarette smoking and lung cancer	80
<b>Baseline reference</b>		
0.1	Smoking cessation before middle age in the UK and lung cancer	90
0.6	HBV vaccination of adults in Korea and liver cancer	40
0.1	HBV vaccination of newborns in Taiwan and liver cancer	90

# Consistenza

- Riproducibilità del risultato in diverse popolazioni o con differenti metodologie investigative

*E questa speranza si faccia più volte, acciò che qualche accidente non impedissi o falsassi tal prova, che la speranza fussi falsa, e ch'ella ingannassi o no il suo speculatore" [Leonardo da Vinci].*

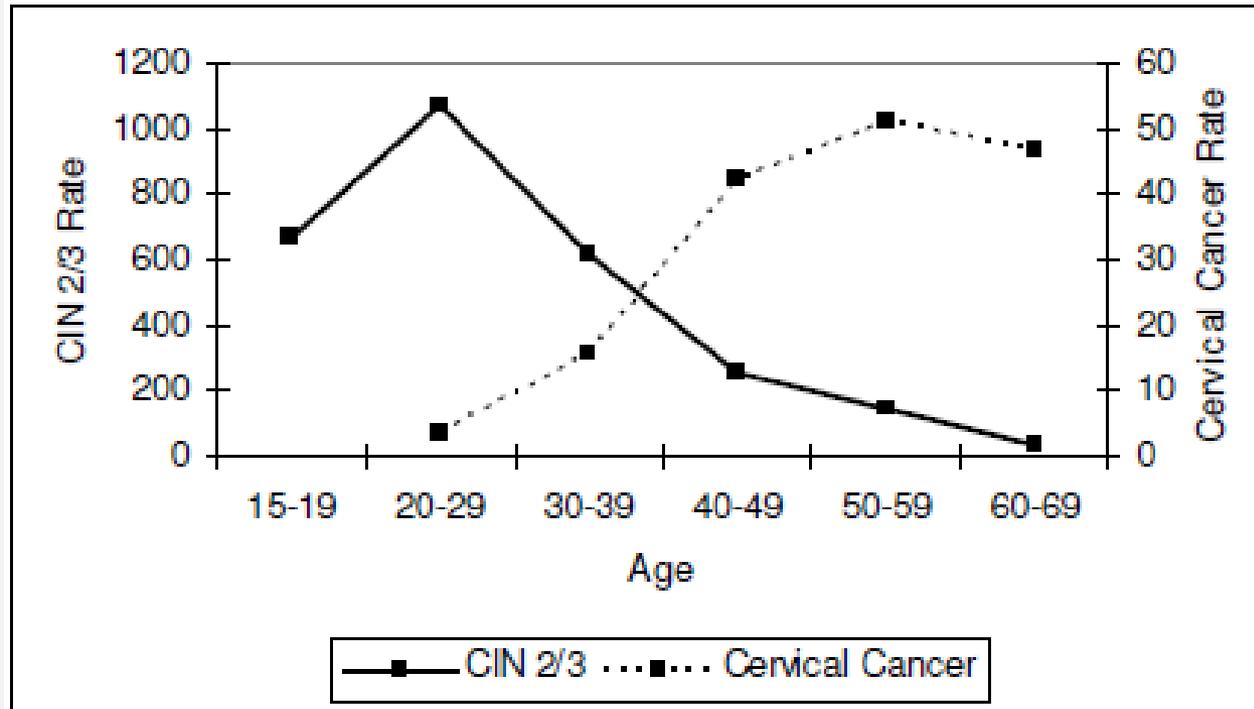
# Prevalenza di HPV DNA in casi di ca. cervicale e in controlli (IARC multicentre study)



# Congruenza temporale

- Il rapporto causale implica l'antecedenza di uno o più fattori rispetto all'effetto.
- Difficile da valutare negli studi caso controllo o trasversali

# Incidenza di CIN2/3 e di cancro cervicale



**Incidenza di CIN 2/3 individuati mediante screening e incidenza di cancro della cervice prima dello screening.** I tassi sono espressi per 100.000 donne sottoposte a screening citologico routinario (CIN2/3) e per 100.000 donne (Ca cervicale). Il picco dell'incidenza di Ca Cervicale si osserva circa 25–30 anni dopo quello di CIN 2/3.

# Specificità

- Criterio minore – predittività del determinante nei confronti dell'effetto

# Plausibilità biologica

- Meno rigoroso, dipende dalle conoscenze del momento.

# Cancri attribuibili ad HPV

Site	Attributable to HPV (%)	Developed countries		Developing countries	
		Total cancers	Attributable to HPV	Total cancers	Attributable to HPV
Cervix	100	83 400	83 400	409 400	409 400
Penis	40	5 200	2 100	21 100	8 400
Vulva, vagina	40	18 300	7 300	21 700	8 700
Anus	90	14 500	13 100	15 900	14 300
Mouth	> = 3	91 200	2 700	183 100	5 500
Oro-pharynx	> = 12	24 400	2 900	27 700	3 300
All cancers	5	5016 100	111 500	5 827 500	449 600

# Criteri di causalità

- Forza dell'associazione: l'associazione tra HPV DNA in campioni cervicali e ca. cervicale è una delle più forti osservate per un cancro umano.
- Consistenza: l'associazione tra HPV DNA in campioni cervicali e ca. cervicale è stata confermata da numerosi studi in paesi e popolazioni diverse; non sono stati pubblicati studi che mettono in dubbio l'associazione.
- Specificità: l'associazione tra tipi specifici di HPV e ca. cervicale non è casuale; tipi istologici di tumore sono associati a tipi specifici di HPV.
- Sequenza temporale: l'infezione da HPV precede le lesioni precancerose e il ca cervicale di un numero congruo di anni.
- Plausibilità biologica: l'associazione tra HPV DNA in campioni cervicali e ca. cervicale è plausibile e coerente con lo stato delle conoscenze. L'azione oncogena di HPV è mediata da interferenze virali con i meccanismi regolatori della crescita cellulare, della riparazione del DNA e della risposta immunitaria.

# Criteri sequenziali di Causalità

Individuazione di possibile determinante



Controllo di bias  
-Di selezione  
-Di misura



Controllo di confounding



Applicazione dei criteri guida

# VALIDITA'

- La validità di uno studio può essere divisa in due componenti:
  - *validità interna*, relativa al campione studiato.
  - *validità esterna*, relativa alla trasferibilità delle inferenze alla popolazione esterna al campione studiato.
  - La validità interna è condizione necessaria, ma non sufficiente, per la validità esterna

# VALIDITA' INTERNA

## Fattori di distorsione

- *bias* di selezione
- fattori confondenti
- *bias* di informazione

# Confounding

- Il confounding può essere considerato come una confusione di effetti.
- L'effetto apparente dell'esposizione al fattore studiato viene distorto perché confuso o mescolato con l'effetto di un fattore estraneo (*confounder*).
- Es: fluorazione delle acque – carie (*igiene orale*)

# Esempio di *confounding*

Consumo di caffè	Mortalità CC in maschi 55-64 (x1000 x anno)
0	6
1-5	8
> 5	12
Consumo sigarette pacchetti	
0	4
1-2	10
> 2	15

# Esempio di *confounding*

		Sigarette (pacchetti/giorno)			
		0	1-2	>2	totale
Tazze di caffè per giorno	0	4	9	15	6
	1-5	6	10	13	8
	> 5	5	9	16	12
totale		4	10	15	

# Criteria per classificare un fattore come *confounder*

- Un fattore confondente deve essere un fattore di rischio per la malattia o un suo marcatore.
- Un fattore confondente deve essere associato con l'esposizione nella popolazione dalla quale sono estratti i casi e non soltanto nei casi stessi.
- Un fattore confondente non deve essere influenzato dall'esposizione o dalla malattia; in particolare non deve essere uno step intermedio nella catena patogenetica tra esposizione e malattia.

# Metodi di controllo del confounding

- Strutturali
  - Restrizione
  - Appaiamento
- Analitici
  - Stratificazione
  - Modelli
  - Analisi multivariata

# Analisi dei dati – mortalità ospedali A e B

		deceduti		
		Sì	No	Totale
ospedale	A	72	1728	1800
	B	96	3504	3600
		168	5232	5400

Epi Info: Analysis of single Table

Relative Risk = 1,50 (1.11<RR<2,03)

Taylor series 95% confidence limits for RR

	Chi-Squares	P-values
Uncorrected	7,08	0,0078065
Yates corrected	6,64	0,0099607

# Analisi dei dati – alta gravità

		deceduti		
		Sì	No	Totale
ospedale	A	69	1281	1350
	B	84	1716	1800
		153	2993	3150

Epi Info: Analysis of single Table

Relative Risk = 1,10 (0,80<RR<1,49)

95% confidence limits for RR

	<b>Chi-Squares</b>	<b>P-values</b>
Uncorrected	0,33	0,566

# Analisi dei dati – bassa gravità

		Infezione ospedaliera		
		Malati	Sani	Totale
Cateterismo	Sì	3	447	450
	No	12	1788	1800
		15	2235	2250

Epi Info: Analysis of single Table

Relative Risk = 1,00 (0,28<RR<3,53)

Taylor series 95% confidence limits for RR

	<b>Chi-Squares</b>	<b>P-values</b>
Uncorrected	0,00	1,0000000
Fisher exact: 1 tailed		0,6026462
2 tailed		1,0000000

# Analisi dei dati – Mantel Haenszel

<b>Tipo</b>	<b>RR</b>	<b>Significato</b>
RR grezzo	1,50 (1,11-2,03)	Associazione statistica
RR alta gravità	1,10 (0,80-1,49)	Nessuna associazione
RR bassa gravità	1,00 (0,28-3,53)	Nessuna associazione
RR Mantel Haenszel	1,09 (0,81-1,47)	Nessuna associazione

# *bias* di selezione

Sono distorsioni che risultano da:

- procedure impiegate per selezionare i soggetti
- fattori che influenzano la partecipazione allo studio

tali che la relazione tra esposizione e malattia è diversa tra chi partecipa e chi non partecipa allo studio.

- *bias di autoselezione*
- *bias diagnostico*

# *Bias* di auto-selezione: esempio

- studiato outcomes del 76% delle truppe presenti al Smoky atomic test in Nevada (8 casi di leucemia)
  - 82% (62% del tot.) individuate dai ricercatori  $\Rightarrow$  4 casi
  - 18% (14% del tot.) autoselezionate  $\Rightarrow$  4 casi
- se per il 24% non studiato l'incidenza fosse quella del primo gruppo dovremmo attenderci ulteriori  $4(24/62) = 1.5$  casi
- se per lo stesso gruppo l'incidenza fosse quella del gruppo complessivo dovremmo attenderci ulteriori  $8(24/76) = 2.5$  casi

# Bias di informazione

- Un bias di informazione si verifica ogni qualvolta vi siano errori nella rilevazione dei dati nei soggetti studiati
- per le variabili discrete prende il nome di errore di classificazione che può essere a sua volta differenziale o non-differenziale in funzione della dipendenza dai valori di altre variabili

# Bias di informazione

- Errore di classificazione differenziale
  - esempio 1: studio di coorte su enfisema e fumo di tabacco; i fumatori vanno dal medico più frequentemente dei non fumatori.
  - Esempio 2: studio caso controllo su malformazioni congenite; le madri di bambini malformati ricordano più accuratamente delle madri dei bambini normali possibili esposizioni ai fattori di rischio (*recall bias*)

# Bias di informazione

- Errore di classificazione non differenziale quando la proporzione dei soggetti classificati erroneamente non dipende dalla presenza o meno della malattia o dell'esposizione
  - *dell'esposizione*
  - *della malattia*

# Errore di classificazione non differenziale dell'esposizione

*Si verifica quando la probabilità di errore di classificazione dell'esposizione è indipendente dallo stato di malattia.*

*Si incontra soprattutto negli studi di coorte.*

# Errore di classificazione non differenziale dell'esposizione

		Malattia		totale
		presente	assente	
Consumo alcool	presente	40	460	500
	assente	20	480	500
totale		60	940	1000

$$RR = 2.00$$

# Errore di classificazione non differenziale dell'esposizione

Sottostima dell'esposizione del 20%

		Malattia		totale
		presente	assente	
Consumo alcool	presente	32	368	400
	assente	28	572	600
totale		60	940	1000

RR = 1.7

# Errore di classificazione non differenziale della malattia

*Si verifica quando la probabilità di erronea classificazione della malattia è indipendente dallo stato di esposizione.*

- Malattie di diagnosi difficile
- Tecniche diagnostiche poco accurate

# Errore di classificazione non differenziale della malattia

		Influenza		totale
		presente	assente	
Vaccino	No	40	460	500
	Sì	20	480	500
totale		60	940	1000

RR = 2.0

# Errore di classificazione non differenziale della malattia

Sensibilità: 100%, specificità: 80%

		Influenza		totale
		presente	assente	
Vaccino	No	132	368	500
	Sì	114	386	500
totale		246	754	1000

RR = 1.16

# Errore di classificazione non differenziale della malattia

Sensibilità Specificità	Incidenza esposti	Incidenza non esposti	Rapporto differenza
100%	40/100	20/200	4.0
100%			0.3
70%	28/100	14/200	4
100%			0.21
100%	52/100	56/200	1.9
80%			0.24