



# Calcolo delle probabilità in ambito sanitario

Prof. Lucio Torelli

Dipartimento di Scienze Mediche, Chirurgiche e della Salute

Università degli Studi di Trieste

Trieste, 20 gennaio 2017

che ci sta a fare un matematico  
a Medicina?

calcolo delle probabilità in ambito  
sanitario?



... mathematics has a **dual nature**: it is both an independent discipline valued for **precision** and intrinsic **beauty**, and it is a rich source of tools for the world of **applications**.

Mathematics might be described as having **abstractness** internally and **effectiveness** externally ...

*Report of the Senior Assessment Panel for the International Panel of the U.S. Mathematical Sciences, 1998*

dal pieghevole del Centro di Ricerca Matematica Ennio De Giorgi – Scuola Normale Superiore di Pisa

- Il problema dell'incertezza
- E' possibile misurare l'incertezza?  
Come siamo abituati a misurare l'incertezza?
- Matematica, incertezza ... calcolo delle probabilità



Diverse definizioni,  
da applicare in maniera “corretta” ...

Definizione **classica**: casi favorevoli / casi possibili  
se “eventi elementari” sono “equiprobabili”

Definizione **frequentista**:  $\cong$  freq. relativa osservata  
se abbiamo un numero “suff. grande” di osservazioni

Definizione **soggettivista**: è il prezzo  $p$  ( $0 \leq p \leq 1$ ) che sono  
disposto a pagare in una scommessa equa ...  
il problema della soggettività ...

**come fare nella pratica in ambito bio-medico?**

# e cos'è la probabilità condizionata?

## Probabilità condizionata (subordinata)

Finora abbiamo parlato di probabilità senza limitazioni, cioè di probabilità **incondizionata (o subordinata)**; però spesso si incontrano eventi che dipendono da altri eventi che si possono (o debbono) verificare precedentemente: tali eventi, naturalmente, influiranno sulla probabilità dell'evento successivo; in tal caso occorre introdurre il concetto di **probabilità condizionata**

**Definiamo probabilità condizionata dell'evento  $E_2$  rispetto all'evento  $E_1$  la probabilità che si verifichi l'evento  $E_2$  dopo che si è verificato l'evento  $E_1$**

$$P(E_2|E_1)$$

Esempio:

trovare la probabilità che la seconda carta estratta da un mazzo di 40 sia una figura se la prima estratta è anch'essa una figura:

in pratica l'uscita della prima figura (evento  $E_1$ ) fa variare la probabilità di uscita della seconda carta (evento  $E_2$ ) perché avremo 11 casi favorevoli su 39 casi possibili

$$P(E_2|E_1) = 11/39$$

Per calcolare la probabilità condizionata possiamo usare la formula

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad \text{fare link per dimostrazione}$$

Vale una formula equivalente per la probabilità condizionata dell'evento  $E_1$  rispetto all'evento  $E_2$

Poiché vale  $E_1 \cap E_2 = E_2 \cap E_1$  allora è valida anche la formula

$$P(E_1|E_2) = \frac{P(E_2 \cap E_1)}{P(E_2)}$$

Esempio:

Trovare la probabilità che, nel lancio di un dado, sapendo che il risultato sarà un numero dispari, si ottenga il numero 1

$E_2|E_1$  = uscita del numero 1 sapendo che esce un numero dispari

$E_1$  = uscita di un numero dispari


$E_2$  = uscita del numero uno

$E_1 \cap E_2$  = uscita del numero 1 e dispari (secondo 1 dispari equivale all'evento uscita del numero 1)

# La probabilità condizionata

$$P(E_2 | E_1) = \frac{P(E_2 \cap E_1)}{P(E_1)}, \quad P(E_1) > 0$$

	Dalt.	Non Dalt.	totali
M	60	540	600
F	5	495	500
totali	65	1035	1100

- $P(\text{Dalt.}) ?$
  - $P(\text{Dalt.} \mid \text{F}) ?$
  - $P(\text{Dalt.} \mid \text{F}) = P(\text{Dalt.}) ??$
- “Sapendo che”*
- 



Siamo in grado di **completare**  
questa tavola di contingenza 2 x 2?

	Malati	Sani	totali
M			60%
F			40%
totali	10%	90%	100%

Siamo in grado di completare  
questa tavola 2 x 2?

	Malati	Sani	totali
M	6%	54%	60%
F	4%	36%	40%
totali	10%	90%	100%

**Se la malattia non dipende dal genere !!!**

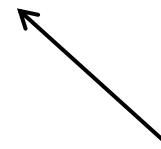
Ad es.  $P(F \cap \text{Sani}) = P(F) * P(\text{Sani}) = 40\% * 90\% = 36\% \dots$

Dalla definizione di probabilità condizionata si può ricavare la **formula di Bayes** ....:

$$P(E | F) = \frac{P(F | E)}{P(F)} * P(E), \quad P(F) > 0$$



Probabilità a posteriori



Probabilità a priori

# i test diagnostici

	T+	T-	
M+			
M-			

**Sensibilità**: è la probabilità che una persona risulti positiva al test, sapendo che è malata

**Specificità**: è la probabilità che una persona risulti negativa al test, sapendo che è sana

$$Se = P(T+ | M+) \cong \frac{T+ \cap M+}{M+} = \frac{\text{veri positivi}}{\text{malati}}$$

$$Sp = P(T- | M-) \cong \frac{T- \cap M-}{M-} = \frac{\text{veri negativi}}{\text{sani}}$$

**Domanda**: In un test con sensibilità del 100% non ci possono essere falsi positivi o falsi negativi?

	T+	T-	
M+			
M-			

**Valore predittivo pos.**: è la probabilità che una persona sia malata, sapendo che è positiva al test

**Valore predittivo neg.**: è la probabilità che una persona sia sana, sapendo che è negativa al test

$$V_{p+} = P(M+|T+) \cong \frac{T+ \cap M+}{T+} = \frac{\text{veri positivi}}{\text{positivi}}$$

$$V_{p-} = P(M-|T-) \cong \frac{T- \cap M-}{T-} = \frac{\text{veri negativi}}{\text{negativi}}$$

**Attenzione a non confondere sensibilità e valore predittivo positivo ...**

# esempio sul pap-test (Papanicolaou)... da Wikipedia:

## **Limiti del Pap test** [[modifica](#)]

*Il Pap test non è indicato per la individuazione dei tumori dell'[endometrio](#) o di altri organi dell'[apparato genitale femminile](#).*

*Per quanto complessivamente il Pap test si sia dimostrato estremamente efficace nel ridurre la frequenza del cancro invasivo del collo dell'utero, come tutte le tecniche di screening presenta dei limiti intrinseci alla metodica. In particolare la **sensibilità** del Pap-test viene valutata in circa **60-70%**. Questo significa che sono possibili falsi negativi, cioè test negativi nonostante la presenza di un tumore, ma anche falsi positivi, cioè casi in cui il risultato positivo del test non viene confermato da successive indagini.*

	T+	T-	
M+			
M-			

sens:	60%
spec:	90%
Prev:	1%

Risulta quindi:

$$Se = P(T+ | M+) = \frac{\text{veri\_positivi}}{\text{malati}}$$

$$Sp = P(T- | M-) = \frac{\text{veri\_negativi}}{\text{sani}}$$

$$\text{veri\_positivi} = Se * \text{malati} = 60\% * 1\% = 0,6\%$$

$$\text{veri\_negativi} = Sp * \text{sani} = 90\% * 99\% = 89,1\%$$

	T+	T-	
M+	0,6%	0,4%	1,0%
M-	9,9%	89,1%	99,0%
	10,5%	89,5%	100%

sens:	60%
spec:	90%
Prev:	1%

Risulta pertanto:

$$Vp+ = P(M+ | T+) = \frac{\text{veri\_positivi}}{\text{test\_positivi}} = 5,7\%$$

$$Vp- = P(M- | T-) = \frac{\text{veri\_negativi}}{\text{test\_negativi}} = 99,6\%$$

Cosa significano tali valori?

Cosa possiamo dire ad una donna T+?



	T+	T-	
M+	12,0%	8,0%	20,0%
M-	8,0%	72,0%	80,0%
	20,0%	80,0%	100%

sens:	60%
spec:	90%
Prev:	20%

Vp+=	60,0%
Vp- =	90,0%

Ma che accade  
se cambia la  
prevalenza??

**Da wikipedia, sett. 2014:**

... **Nel prossimo futuro**, il ruolo del pap test nella prevenzione dei tumori del collo uterino è sicuramente destinato a cambiare. La scoperta che la maggior parte dei tumori del collo uterino sono dovuti al virus del papilloma umano (HPV) ha portato allo sviluppo di tecniche diagnostiche biomolecolari caratterizzate da una **sensibilità elevata (superiore al 95%)** che ne ha fatto prospettare l'utilizzazione come metodica di screening. Ancora discusso è tuttavia il problema della relativa **specificità** delle tecniche biomolecolari di tipizzazione dell'HPV. L'infezione da HPV è infatti largamente diffusa, ed è evidenziabile anche in molte donne in cui tuttavia lo HPV è solo transitorio e non è destinato a causare lo sviluppo di un tumore. ...

	T+	T-	
M+	0,95%	0,05%	1,0%
M-	9,90%	89,10%	99,0%
	10,85%	89,15%	100%

sens:	95%
spec:	90%
Prev:	1%

Risulta pertanto:

$$Vp_+ = P(M_+ | T_+) = \frac{\text{veri\_positivi}}{\text{test\_positivi}} = 8,8\%$$

$$Vp_- = P(M_- | T_-) = \frac{\text{veri\_negativi}}{\text{test\_negativi}} = 99,9\%$$

	T+	T-	
M+	19%	1%	20%
M-	8%	72%	80%
	27%	73%	100%

sens:	95%
spec:	90%
Prev:	20%

Vp+=	70,4%
Vp-:	98,6%

# Le misure del rischio

(in coll. con la dott.ssa Giudici)

# Prevention of Breast Cancer in Postmenopausal Women: Approaches to Estimating and Reducing Risk

Steven R. Cummings, Jeffrey A. Tice, Scott Bauer, Warren S. Browner, Jack Cuzick, Elad Ziv, Victor Vogel, John Shepherd, Celine Vachon, Rebecca Smith-Bindman, Karla Kerlikowske

## Endogenous Hormone Levels and Estimation of Breast Cancer Risk

A study (34) that combined data from nine prospective cohort studies of breast cancer found that women in the highest quintile of estradiol or testosterone had a higher relative risk of breast cancer (for estradiol, 2.00-fold, 95% CI = 1.47- to 2.71-fold; for testosterone, 2.22-fold, 95% CI = 1.59- to 3.10-fold) than those in the lowest quintile. We found six subsequent prospective studies (32,35–39), all of which found statistically significant associations between estradiol or testosterone levels and the risk of breast can-

Che differenza c'è tra uno studio di coorte o uno studio caso controllo??

Perché alcune volte negli articoli troviamo scritto

Rischio relativo (RR) e altre odds ratio (OR)???

# Injectable and Oral Contraceptive Use and Cancers of the Breast, Cervix, Ovary, and Endometrium in Black South African Women: Case–Control Study

*Methods and Findings:* We analysed data from a South African hospital-based case-control study of black females age 18–79 y, comparing self-reported contraceptive use in patients with breast ( $n = 1,664$ ), cervical ( $n = 2,182$ ), ovarian ( $n = 182$ ) and endometrial ( $n = 182$ ) cancer, with self-reported contraceptive use in 1,492 control patients diagnosed with cancers with no known relationship to hormonal contraceptive use. We adjusted for potential confounding factors, including age, calendar year of diagnosis, education, smoking, alcohol, parity/age at first birth, and number of sexual partners. Among controls, 26% had used injectable and 20% had used oral contraceptives. For current and more recent users versus never users of oral or injectable contraceptives the odds ratios (ORs) for breast cancer were significantly increased in users of oral and/or injectable contraceptives (OR 1.66, 95% CI 1.28–2.16,  $p < 0.001$ ) and separately among those exclusively using oral (1.57, 1.03–2.40,  $p = 0.04$ ) and exclusively using injectable (OR 1.83, 1.31–2.55,  $p < 0.001$ ) contraceptives; corresponding OR for cervical cancer were 1.38 (1.08–1.77,  $p = 0.01$ ), 1.01 (0.66–1.56,  $p = 0.96$ ), and 1.58 (1.16–2.15,  $p = 0.004$ ). There was no

# Fattore di rischio

- Esposizioni, comportamenti o attributi personali che influenzano il rischio di sviluppare una malattia
- Fattori che aumentano la probabilità di comparsa di una malattia o di una determinata condizione
- 4 categorie principali:
  - Componenti genetiche o ereditarie
  - Componenti ambientali
  - Fattori nocivi presenti nell'ambiente di lavoro
  - Abitudini tipicamente individuali

# Associazione tra “esposizione” e malattia

- l'esistenza della associazione si dimostra confrontando la **frequenza della malattia** in due popolazioni costituite rispettivamente da persone **esposti** e **non esposti** ad una presunta causa (o determinante o «fattore di rischio»)
- **Per “esposizione”** si intende la presenza di qualsiasi variabile che, in linea di ipotesi, può causare un certo effetto



# **DUE APPROCCI PER LA SCOPERTA DI UN'ASSOCIAZIONE**

**STUDIO PROSPETTICO o di COORTE**

**STUDIO RETROSPETTIVO o CASO-CONTROLLO**

# Studi Prospettivi

## Disegno di uno Studio Prospettivo

Reclutamento due gruppi di **individui sani**

- uno composto da soggetti esposti al fattore di rischio
- uno composto da soggetti quanto più possibile simili ai primi, salvo per il fatto di non essere esposti

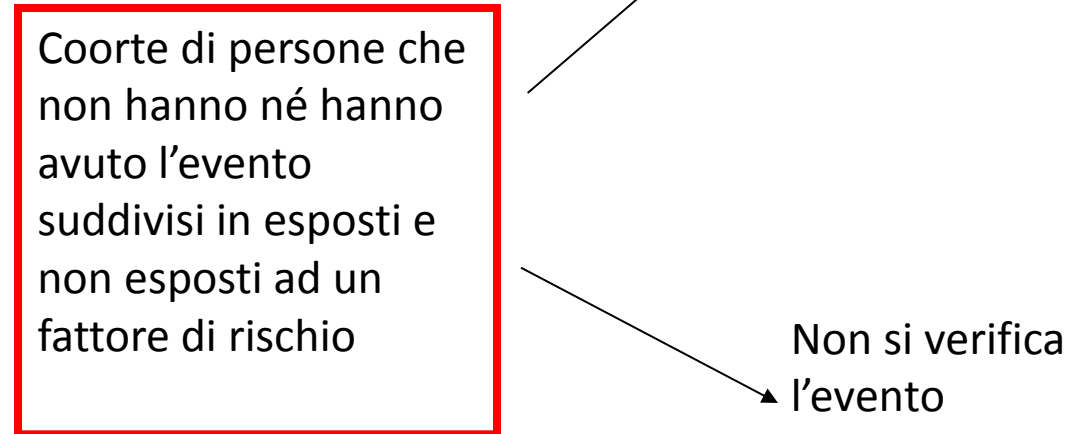
Osservazione dei singoli individui per un periodo di tempo appropriato (follow-up) al fine di **stimare l'incidenza** di malattia nei due gruppi di esposizione (esposti e non esposti).

# Studio prospettico

Passato

Presente

Futuro



In uno studio prospettico, all'inizio l'evento "malattia" non è ancora avvenuto: trascorso un periodo di tempo si determina quanti individui hanno sviluppato la malattia tra gli esposti o non esposti ad un fattore di rischio

# CALCOLO DEL RISCHIO RELATIVO (RR)

	Ammalati (M+)	Non ammalati(M-)	Totale
Esposti	a	b	a+b
Non esposti	c	d	c+d

Numericamente l'associazione tra "fattore di rischio e malattia" è misurata come rapporto (**Rischio Relativo, RR**) tra tassi di incidenza, ovvero come rapporto tra il rischi negli esposti e quello nei non esposti .

**RR=**

**Rischio negli esposti:  $a/(a+b)$**

**Rischio nei non esposti:  $c/(c+d)$**

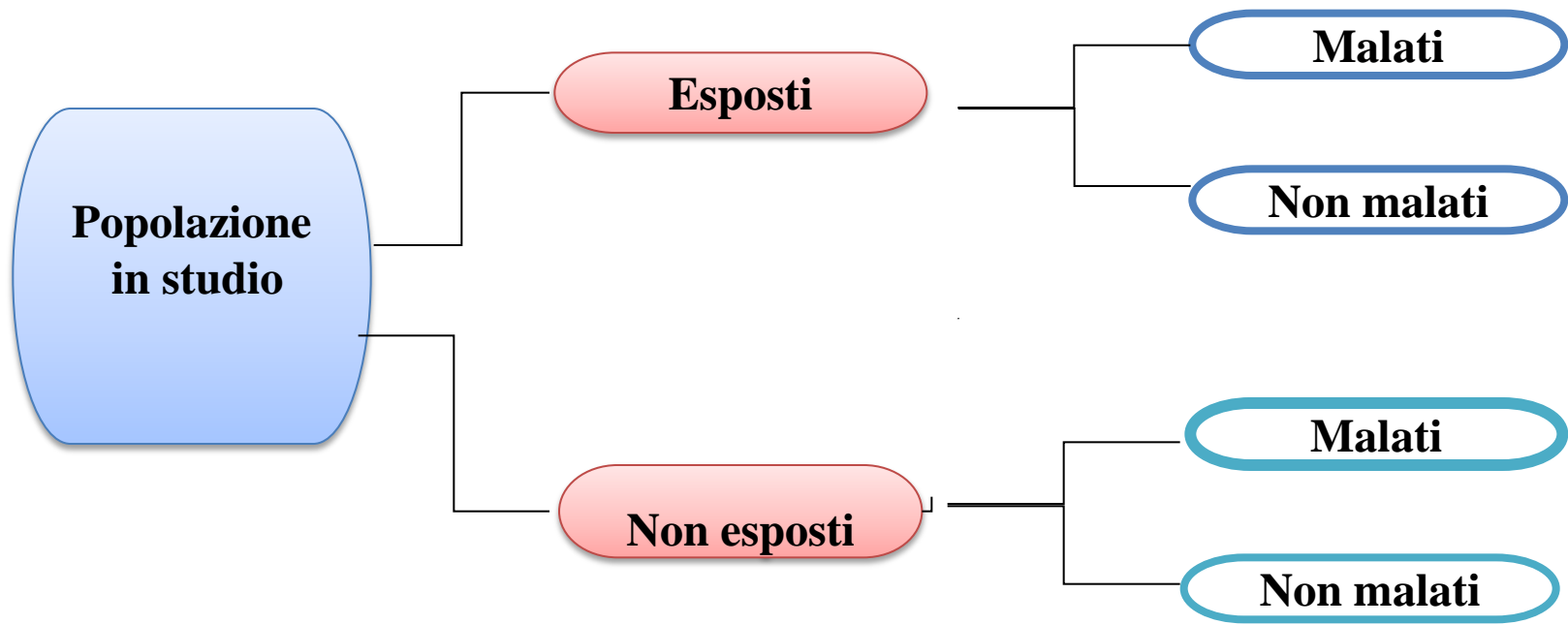
## Troppa carne rossa nella dieta?

**Rischio del 32% in più** di tumore del seno positivo per i recettori degli estrogeni e del progesterone nelle donne dai 26 ai 46 anni: è la conclusione di uno studio pubblicato dagli *Archives of Internal Medicine* che sta attirando l'attenzione degli oncologi di tutto il mondo.



I ricercatori del Department of Medicine del Brigham and Women's Hospital e della **Harvard Medical School** hanno analizzato i dati riguardanti **90.659 donne** in fase premenopausale dai 26 ai 46 anni di età.

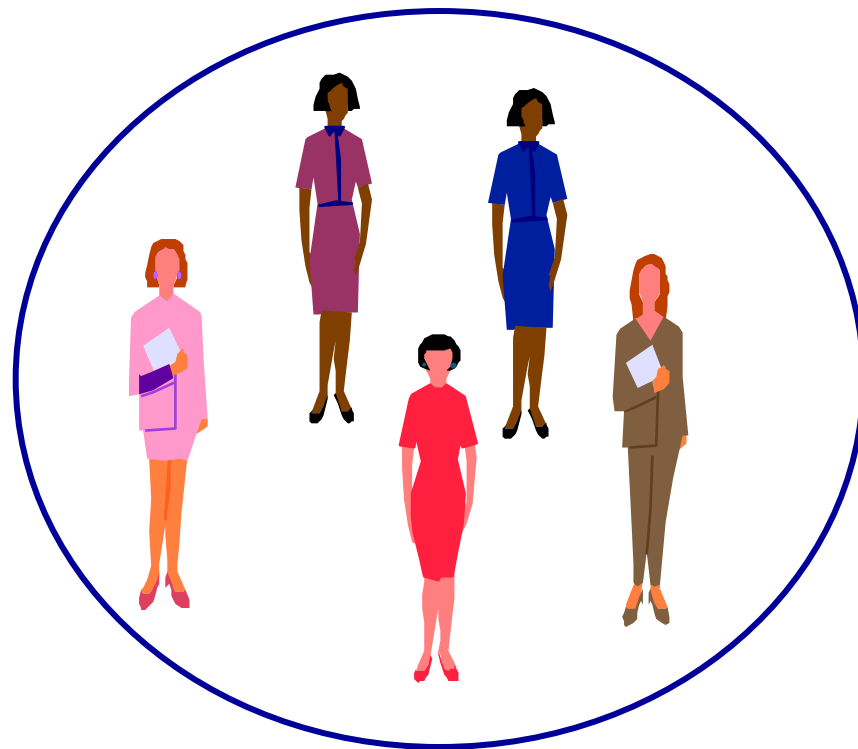
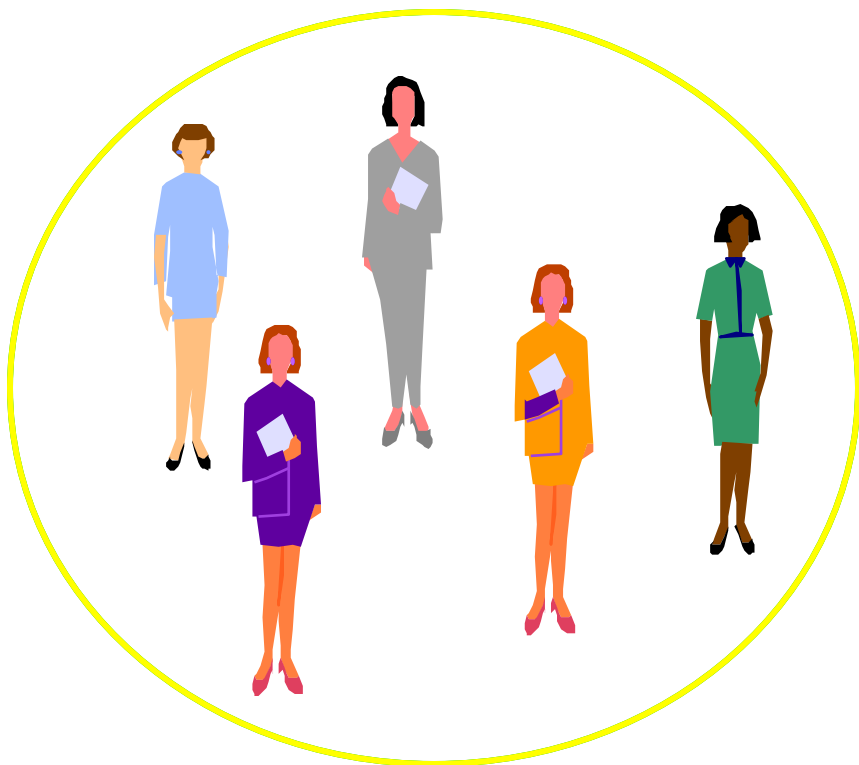
Il **consumo di carne rossa** è stato misurato mediante **questionari** sui regimi alimentari sottoposti alle pazienti ogni 4 anni nell'arco di 12 anni: i **casi** registrati di carcinoma invasivo della mammella sono stati **1021**.



Follow-up di 12 anni

## Esposte

(carne rossa >1,5 al giorno)= 8649 donne



## non esposte

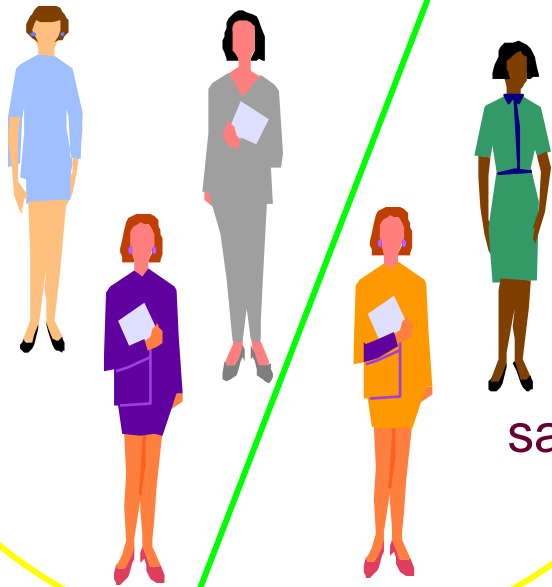
(carne rossa  $\leq$  1,5 volte alla settimana)= 26787 donne



In ciascuno di questi due gruppi si sono presentati casi di malattia con una certa frequenza.

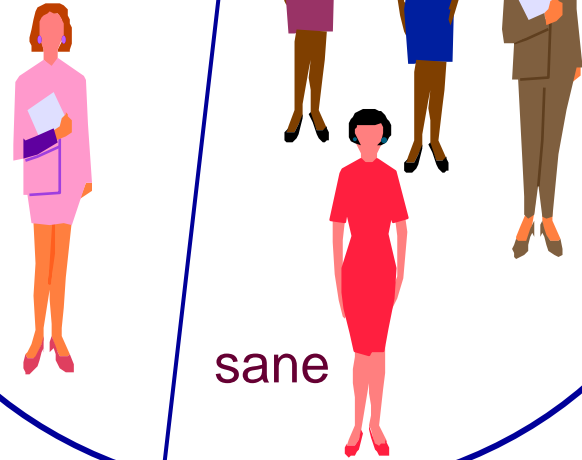
esposte

Malate=52



sane

Malate=121



sane

non esposte

Ipotizziamo a questo punto di aver concluso il nostro studio di coorte e di voler misurare l'eventuale associazione esistente

	Malate	Non Malate	
Esposte	52	8597	8649
Non Esposte	121	26666	26787
Totale	173	35263	35436

$$RR = \frac{\text{rischio nelle esposte}}{\text{rischio nelle non esposte}} = \frac{52/8649}{121/26787} = 1.32$$

Il valore di RR ottenuto (a patto che sia soddisfatta la condizione iniziale di fondamentale uniformità fra i due gruppi confrontati) ci informa su quanto aumenta la probabilità di contrarre la malattia se ci si espone al fattore di rischio:

come di interpreta il risultato ottenuto???

RISCHIO RELATIVO (RR) = misura della forza dell'associazione  
tra fattore di rischio e malattia

$$RR < 1$$

il fattore protegge dal rischio di insorgenza della  
malattia

$$RR = 1$$

il fattore non influenza l'insorgenza della  
malattia

$$RR > 1$$

il fattore aumenta il rischio di insorgenza della  
malattia

# Interpretare il RR

Schema di interpretazione del rischio relativo e dell'odds ratio.



Il RR può essere interpretato anche come eccesso di rischio (EER excess relative risk) rispetto ad un valore di riferimento che per definizione è 1 (RR-1) e per comodità si esprime in %

# Interpretazione delle misure

- Se  $RR = 5$      $5-1=4 \rightarrow 400\%$     Le persone esposte hanno un eccesso di rischio del 400%
  - Le persone esposte è cinque volte più verosimile che abbiano la malattia rispetto ai non esposti
- Se  $RR = 0.5$      $0,5-1=-0,5 \rightarrow -50\%$     Le persone esposte hanno un rischio minore del 50%
  - Le persone non esposte è due volte più verosimile che abbiano la malattia rispetto agli esposti
- Se  $RR = 1$      $1-1=0 \rightarrow 0\%$     Le persone esposte hanno un rischio dell'0%
  - Non c'è associazione tra esposizione e malattia

# Studi retrospettivi

- Gli studi retrospettivi (caso-controllo) sono basati sempre sul reclutamento di due gruppi di individui da una stessa popolazione: uno di soggetti che hanno contratto la malattia in studio (**casi**) ed uno di individui che non l'hanno contratta (**controlli**) ma con caratteristiche simili ai casi
- Si valuta **retrospettivamente la precedente esposizione** al fattore di rischio , sia nei casi che nei controlli: su questi individui vengono raccolte informazioni circa l'esposizione pregressa al fattore di rischio (questionari, interviste).

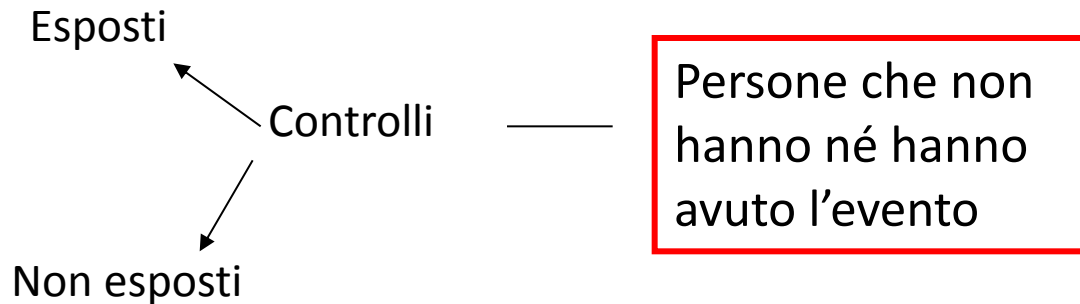
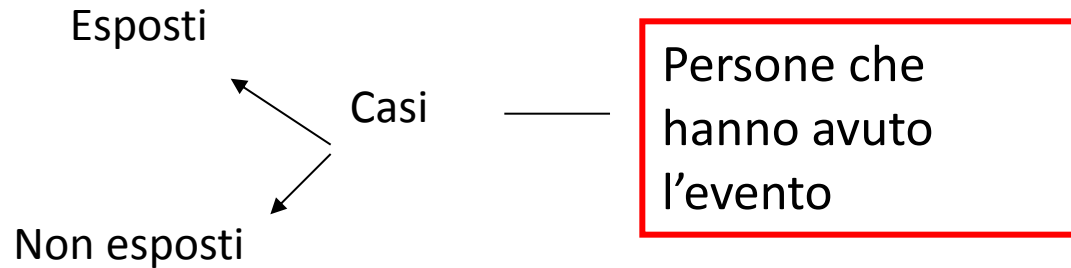
# Caso-controllo: studio retrospettivo

Passato

Presente

Futuro

---



Anche negli studi retrospettivi, i risultati possono essere riassunti in una tabella 2x2 **ma non si può** stimare l'associazione tra malattia e esposizione attraverso il rapporto tra la presenza di malattia negli esposti e nei non esposti, perché il numero di casi e controlli è scelto dal ricercatore

L'associazione tra fattore di rischio e malattia è misurata come rapporto tra **odds** di esposizione per casi e odds di esposizione per i controlli (ODDS RATIO) ....  
ma cosa si intende con odds???



# ...l'odds...

La parola odds deriva dal mondo delle scommesse (classicamente anglosassone...) e non è facilmente traducibile in italiano.

Dal vocabolario

ODDS = probabilità, quotazione



**L'odds, in termini matematici, rappresenta pertanto il rapporto fra eventi positivi ed eventi negativi.**

**L'odds ratio (OR) è un rapporto tra odds**

## CALCOLO DELL'ODDS RATIO (OR)

	<b>Ammalati (Casi)</b>	<b>Non ammalati (Controlli)</b>	<b>Totale</b>
<b>Esposti</b>	<b>a</b>	<b>b</b>	<b>a+b</b>
<b>Non esposti</b>	<b>c</b>	<b>d</b>	<b>c+d</b>
<b>Totale</b>	<b>a+c</b>	<b>b+d</b>	

Odds di esposizione nei casi =  $a/c$

Odds di esposizione nei controlli =  $b/d$

$$OR = \frac{\textit{odds di esposizione nei casi}}{\textit{odds di esposizione nei controlli}} = \frac{a/c}{b/d} = \frac{a * d}{b * c}$$

# Esempio

Viene condotto uno studio riguardante l'associazione tra obesità (esposizione) e urolitiasi (patologia): si selezionano 1014 casi e 1487 controlli e si va a valutare in entrambi i gruppi il fattore "obesita"

	Urolitiasi si Casi	Urolitiasi no Controlli	Totale
Obesità si	383	322	705
Obesità no	631	1165	1796
Totale	1014	1487	2501

$$\text{Odds casi} = a/c = 383/631 = 0,607$$

$$\text{Odds controlli} = b/d = 322/1165 = 0,277$$

$$\text{Odds ratio} = 0,607 / 0,277 = 2,19$$