CDL in SCIENZE RIABILITATIVE DELLE PROFESSIONI SANITARIE

STATISTICA PER LA RICERCA

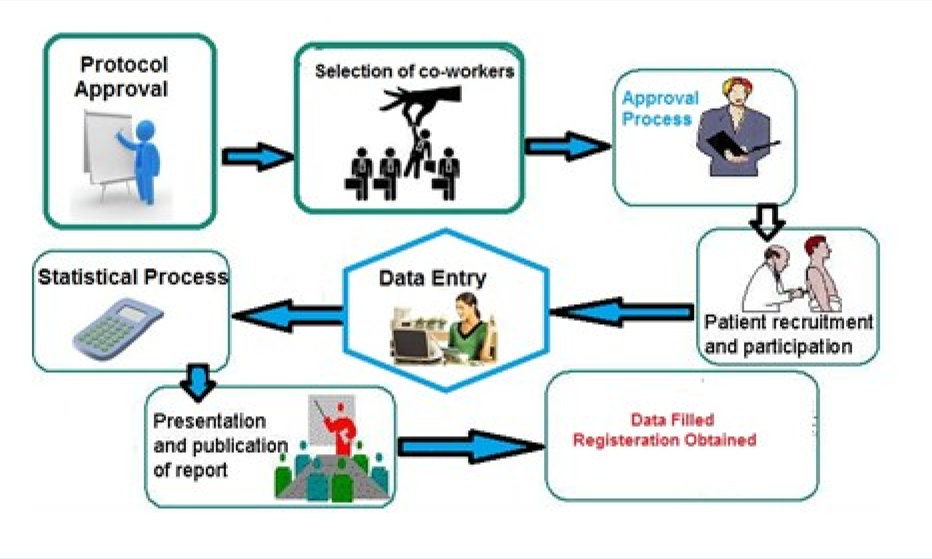
gbarbati@units.it

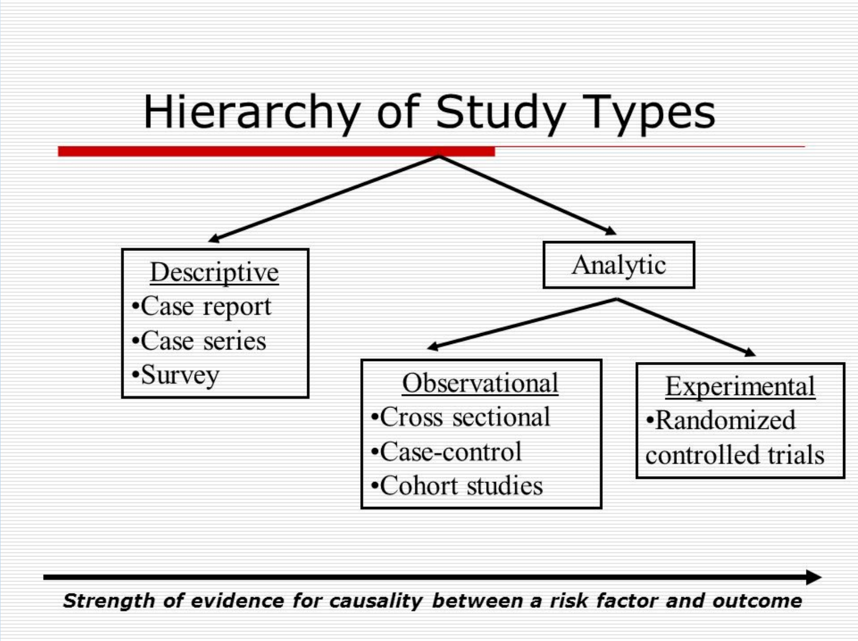
A.A. 2025-26

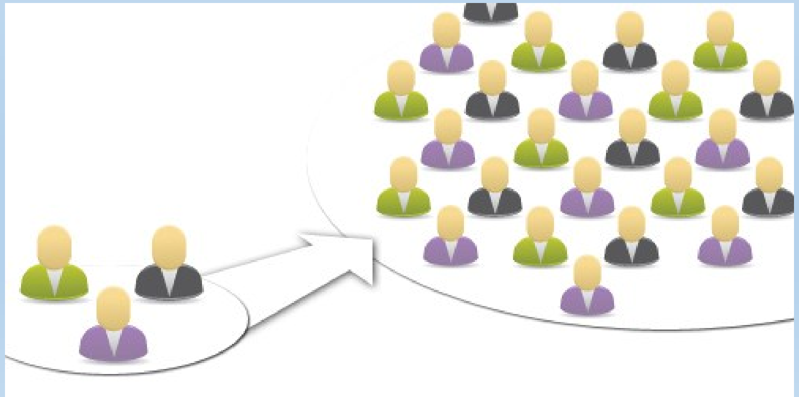
UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Sommario

- Disegni di Studio (recap!)
- Misure di Rischio







Tipi di studi & Obiettivi

Negli studi osservazionali **è più complesso rispetto agli RCTs** definire **a priori** una specifica ipotesi di studio

Varie complicazioni aggiuntive:
(**assenza di randomizzazione**)

- Confondenti (# covariate…)
- «Perdita» al follow-up…
- Valori mancanti….

Obiettivo dello Studio

*Stima di una Prevalenza (**p**)*

Stima di una «media» (μ)

Odds ratio (OR)

Rischio relativo/tempo ad un evento (**RR**/**HR**)

Impatto di parametri diversi sull'outcome: modelli di regressione (β)

Measures of Disease-Exposure Association

Purposes of epidemiology:

- **quantification** of the occurrence of a disease [***descriptive*** studies]
- **strength of the association*** between exposure and the outcome [***analytical*** studies]

Estimate measures of *disease-exposure* association (or **measures of effect**).

Disease frequency in the **exposed group** is compared with the frequency of disease in the **group of those not exposed**, making use of the appropriate measure of occurrence.

This comparison can occur in two ways: in **absolute** terms and in **relative** terms.

***note that we are not using the term *causal* effect!**

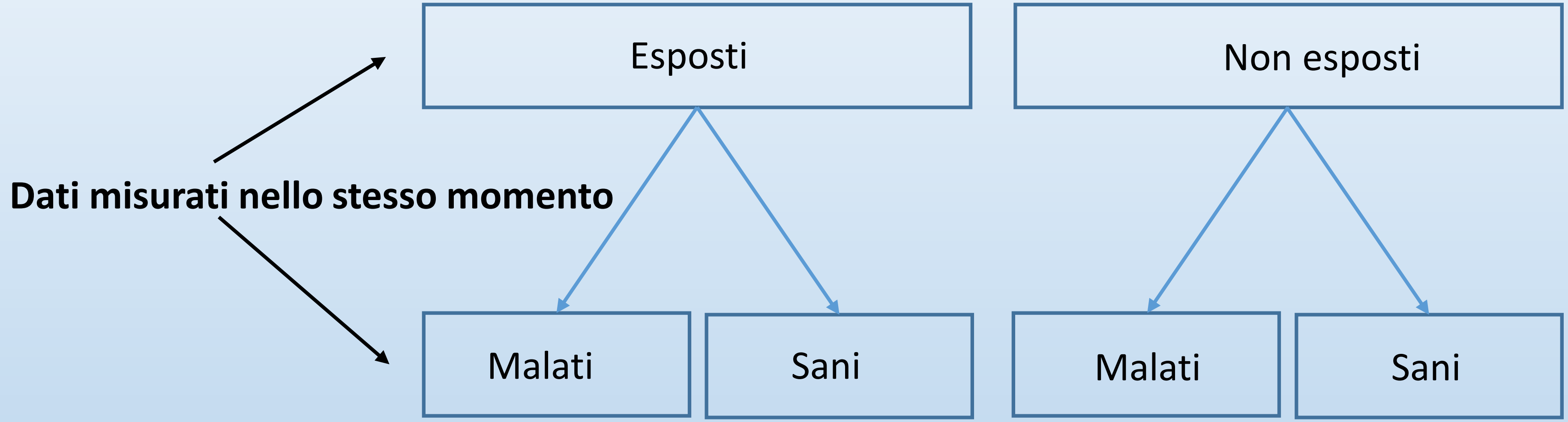
In a *general* sense, each disease could be the *effect* of one (or more…) factors.

In a *quantitative* sense, an **effect** is the measure of ***diversity* in the occurrence** of a disease in two [or more..] groups that differ by *one certain* feature [*univariable* analysis].

- absolute scale: **difference** between two prevalences, two risks (Cum Inc) or two incidence rates
- **relative scale**: **ratio** of two prevalences, two risks (Cum Inc) or two incidence rates

Studi Cross-sectional/trasversali

Studio osservazionale in cui l'esposizione e l'outcome sono determinati **simultaneamente** per ogni soggetto.



Valutano prevalenza malattia / prevalenza esposizione; test di associazione (Chi-quadro)

Utili per:

- screening delle ipotesi sul quesito in studio, rapidi da realizzare;
- valutazioni di prevalenza;
- valutazioni di fattori associati all'outcome di interesse.

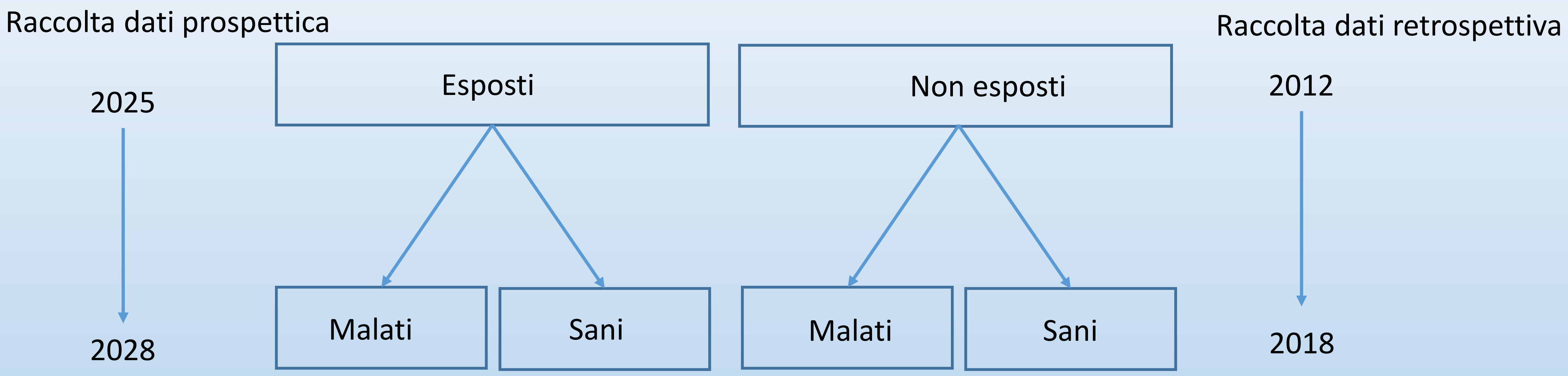
1. Soggetti campionati dalla popolazione target
2. Misura della esposizione & outcome

Limiti:

- Non c'è evidenza temporale di causa-effetto tra l'esposizione e l'outcome;
- Outcomes **prevalenti** ma non incidenti, esclusi i soggetti deceduti prima dello studio (bias «long survivors»);
- A causa della presenza di possibili confondenti bisogna valutare spiegazioni alternative alle associazioni osservate

Studi di coorte

1. Soggetti campionati dalla popolazione di esposti e di non esposti
2. Misura dell'outcome



Utili quando:
- C'è evidenza di una associazione temporale tra esposizione ed outcome;
- L'intervallo temporale tra esposizione e outcome non è troppo lungo;
- L'outcome non è troppo raro.

Limiti:
- Perdita al follow up;
- A causa della presenza di possibili confondenti bisogna valutare spiegazioni alternative
- Bias nella determinazione dell'outcome: non sono blinded!

Valutano **incidenza/rischi relativi** [e assoluti]

Example: does a mother's marital status affect the risk of a baby's death in the first year? ***To what extent***?
What about birthweight?

Relative risk

The Relative Risk for an outcome D associated with a binary risk factor E, denoted by RR, is defined as follows:

$$RR = \frac{P(D|E)}{P(D|\bar{E})}$$

RR=(a/a+b):(c/c+d)

	D	Not D	Tot
E	a	b	a+b
Not E	c	d	c+d
Tot	a+c	b+d	N

Some simple implications immediately follow:

RR

<1 : lower risk or probability of *D* when exposed than when unexposed

=1 : *null* value equivalent to saying that *D* and *E* are independent $P(D|E) = P(D|\bar{E})$

>1 : greater risk or probability of *D* when exposed than when unexposed

The Relative Risk is the basis of a *multiplicative model for risk* :

risk of disease[exposed] = [baseline (unexposed) risk] * RR.

If you smoke cigarettes, your lifetime risk of lung cancer increases tenfold, i.e., the Relative Risk for lung cancer associated with cigarette smoking is 10.

Restrictions on the range:

$$0 < RR \leq \frac{1}{P(D|\bar{E})}$$

For instance, if $P(D|\bar{E}) = 1/3$ then RR $\leq$ 3 since $P(D|E) \leq$1

This restriction only becomes an issue with common disease outcomes

RR is *not symmetric* in the role of the two factors D and E.

The Relative Risk for E associated with D is a different measure of association:

$$\frac{P(D|E)}{P(D|\bar{E})} \neq \frac{P(E|D)}{P(E|\bar{D})}$$

Infant Mortality	Mother's Marital Status: Unmarried	Married
Death	16,712	18,784
Live at 1 year	1,197,142	2,878,421
Total	1,213,854	2,897,205

The Relative Risk for infant mortality in the U.S. in 1991, associated with a mother being unmarried at the time of birth, is:

$$RR = \frac{16712}{1213854} : \frac{18784}{2897205} = 2.12$$

the risk of an infant death with an unmarried mother is **double** the risk w.r.t. mother is married.

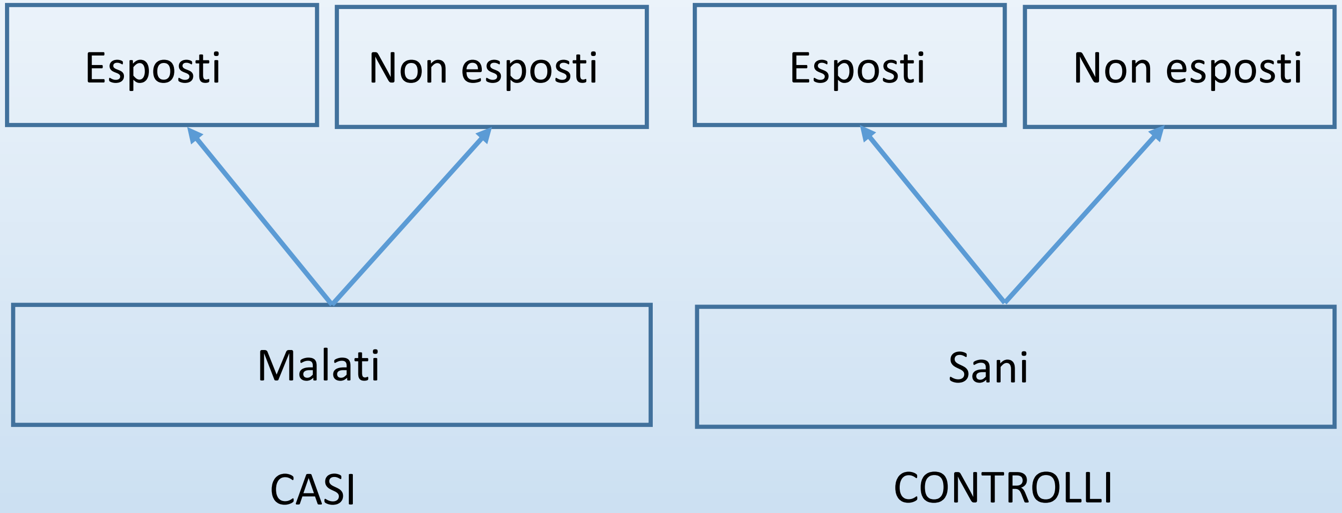
The RR for infant mortality in the U.S. in 1991, associated with a low-birthweight infant, is:

	Birthweight		
Infant Mortality	Low Birthweight	Normal Birthweight	Total
Death	21,054	14,442	35,496
Live at 1 year	271,269	3,804,294	4,075,563
Total	292,323	3,818,736	4,111,059

$$RR = \frac{21054}{292323} : \frac{14442}{3818736} = 19.0$$

Much greater effect of birthweight on infant mortality than we saw for a mother's marital status.

Studi caso-controllo



Gli studi caso-controllo selezionano dei soggetti con l'outcome di interesse (CASI) e li confrontano a soggetti senza (CONTROLLI)

Utili quando:
- L'outcome è raro;
- C'è una ragionevole evidenza di una esposizione nel passato;

Valutano **odds ratios** e non rischi assoluti

Limiti:
- «Recall/Interviewer bias»: i casi possono essere più precisi nel riferire sulla (passata) esposizione;
- Difficile la selezione di controlli adeguati;
- Se si utilizzano procedure di **«matching»** da valutare la generalizzabilità dei risultati;
- Confondenti : bisogna valutare spiegazioni alternative alle associazioni osservate

Odds ratio

An alternative quantity that is used in evaluating risks is the ***odds*** of D as given by: $\frac{P(D)}{P(\bar{D})}$

The odds gives the likelihood of D occurring relative to it not occurring: “how likely am I to win?” as compared to “how likely am I to lose?”

An even odds event D (odds of D are 1) is equivalent to P(D)=1/2, that is, the **same chance** of winning as losing.

The Odds Ratio measures association by comparing the odds of D in the exposed and unexposed.

The Odds Ratio for D associated with E is defined by:

$$OR = \frac{P(D|E)}{P(\bar{D}|E)} : \frac{P(D|\bar{E})}{P(\bar{D}|\bar{E})}$$

OR=(a/c):(b/d)

	Malati	**Sani**
Esposti	**a**	**b**
Non Esposti	**c**	**d**

OR

<1 : lower risk or probability of *D* when exposed

$=1$: *null* value equivalent to saying that *D* and *E* are independent

>1 : greater risk or probability of *D* when exposed

The Odds Ratio is also the basis of a *multiplicative model* for the risk of D.

Like RR, OR > 0, but unlike RR, OR *has no upper limit* whatever the baseline risk $P(D|\bar{E})$ is.

Thus, the OR can be effectively used as a scale for association even when $P(D|\bar{E})$ is large.

	Mother's Marital Status	
Infant Mortality	Unmarried	Married
Death	16,712	18,784
Live at 1 year	1,197,142	2,878,421
Total	1,213,854	2,897,205

The OR for infant mortality associated with an unmarried mother is:

$$OR = \left[\frac{16712}{1213854} : \frac{1197142}{1213854}\right] : \left[\frac{18784}{2897205} : \frac{2878421}{2897205}\right] = 2.14$$

Associated with low birthweight, the OR is:

$$OR = \left[\frac{21054}{292323} : \frac{271269}{292323}\right] : \left[\frac{14442}{3818736} : \frac{3804294}{3818736}\right] = 20.4$$

	Birthweight		
Infant Mortality	Low Birthweight	Normal Birthweight	Total
Death	21,054	14,442	35,496
Live at 1 year	271,269	3,804,294	4,075,563
Total	292,323	3,818,736	4,111,059

Symmetry of roles of disease and exposure in the odds ratio

The Odds Ratio is notoriously confusing when first encountered, particularly in contrast to the simplicity of the interpretation for the Relative Risk. Why is the Odds Ratio then used so often*? A fundamental reason is that the Odds Ratio is ***symmetric*** in the roles of D and E.

Reversing the roles of D and E makes **no difference** in Odds Ratio : this is the **key** to estimating association between an exposure and disease in **case-control studies**.

$$\boxed{OR = \frac{P(D|E)}{P(\bar{D}|E)} : \frac{P(D|\bar{E})}{P(\bar{D}|\bar{E})}} = \frac{P\,(D\&E)/P\,(E)}{P\,(\bar{D}\&E)/P\,(E)} : \frac{P\,(D\&\bar{E})/P\,(\bar{E})}{P\,(\bar{D}\&\bar{E})/P\,(\bar{E})}$$

$$= \frac{P(D\&E)}{P(\bar{D}\&E)} : \frac{P(D\&\bar{E})}{P(\bar{D}\&\bar{E})} = \frac{P(D\&E)}{P(D\&\bar{E})} : \frac{P(\bar{D}\&E)}{P(\bar{D}\&\bar{E})}$$

$$= \frac{P\,(D\&E)/P\,(D)}{P\,(D\&\bar{E})/P\,(D)} : \frac{P\,(\bar{D}\&E)/P\,(\bar{D})}{P\,(\bar{D}\&\bar{E})/P\,(\bar{D})}$$

$$\boxed{= \frac{P(E|D)}{P(\bar{E}|D)} : \frac{P(E|\bar{D})}{P(\bar{E}|\bar{D})}}$$

*Also for the popularity of the **logistic regression** model

The odds ratio as an approximation to the relative risk

If the risk of disease is **low** - that is, the disease is **rare** - in both exposed (E) and unexposed (not E), $P(\bar{D}|E)$ and $P(\bar{D}|\bar{E})$ are both close to 1 and the OR and the RR are *approximately* equal:

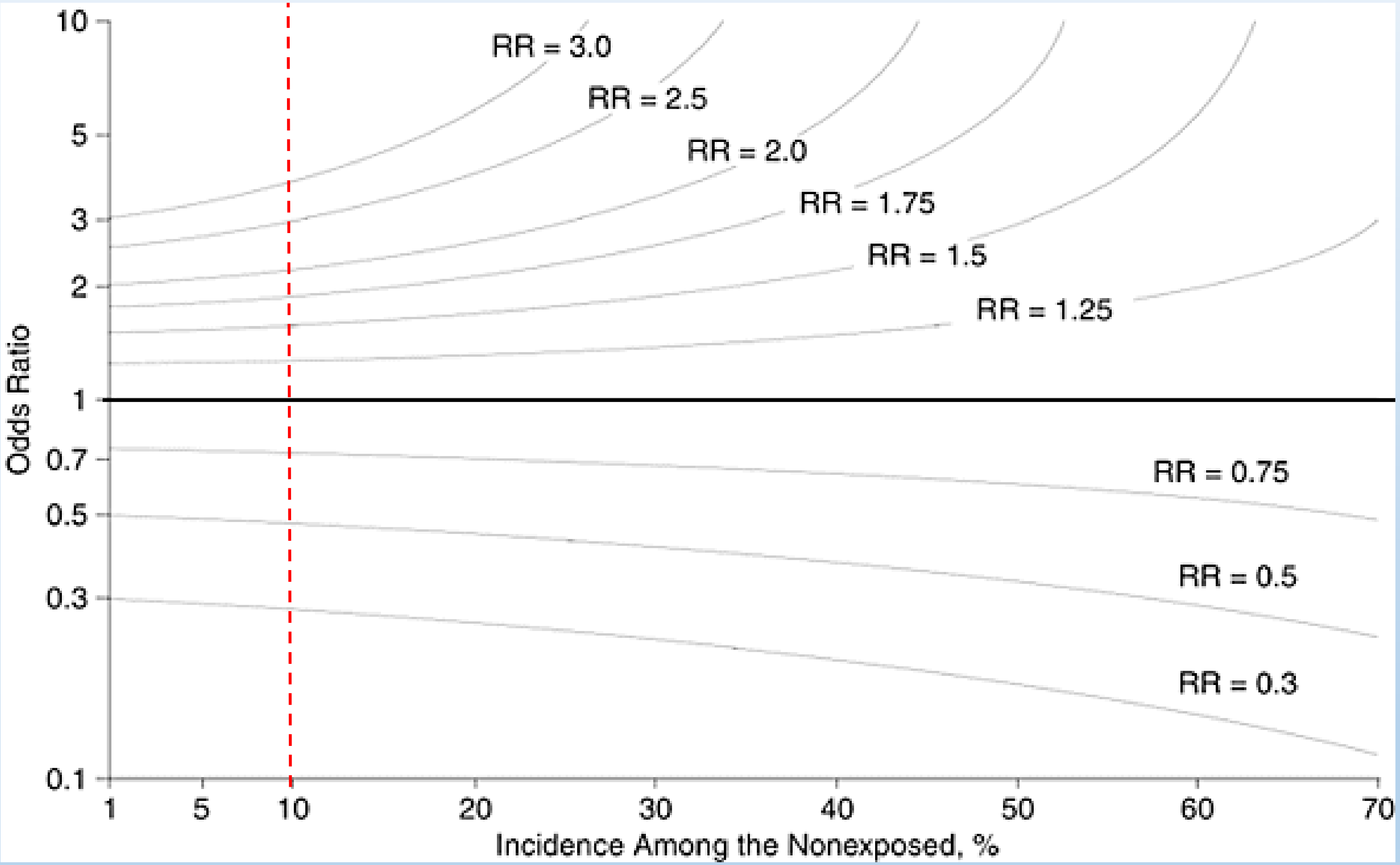
$$P(\bar{D}|E) \approx P(\bar{D}|\bar{E}) \approx 1$$

$$OR \approx \frac{P(D|E)}{P(D|\bar{E})} = RR$$

Generally, OR is similar to the RR when the *sum of the risks* - in the exposed and unexposed - is < 0.1

https://jamanetwork.com/journals/jama/fullarticle/188182

The relationship between relative risk (RR) and odds ratio (OR) by incidence of the outcome:



When the ***incidence*** of an outcome is low (<10%), the odds ratio is *close* to the relative risk.

The more frequent the outcome becomes, the more the odds ratio will **overestimate** the relative risk when it is more than 1 or **underestimate** the relative risk when it is less than 1.

https://www.youtube.com/watch?v=76W4Wymv2Ec

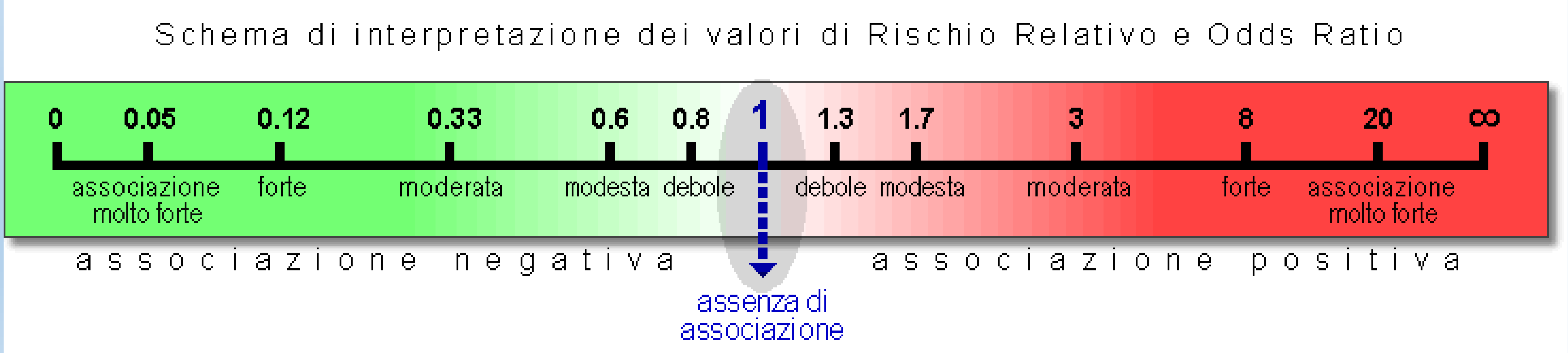
L'interpretazione dei valori numerici è analoga sia che si tratti di valori di OR che di valori di RR.

È intuitivo che un valore **=1** indica **assenza di associazione** tra malattia ed esposizione, in quanto:

- per il **rischio relativo**: l'incidenza negli esposti è uguale all'incidenza nei non esposti;
- per **l'odds ratio**: gli odds di esposizione nei casi sono uguali agli odds di esposizione nei controlli.

Un valore **<1** indica una associazione negativa (cioè il fattore può ***proteggere*** dalla malattia) mentre un rapporto **>1** indica l'esistenza di una associazione positiva (il fattore può ***aumentare il rischio*** della malattia).

Più i valori si discostano da 1, *in un senso o nell'altro…*, più l'associazione è forte.



Misure di Rischio: calcolo dell'intervallo di confidenza per RR

	Apgar score <7	Apgar score >7	Tot
Crescita fetale simmetrica	2 (a)	14 (b)	16 (a+b)
Crescita fetale asimmetrica	33 (c)	58 (d)	91 (c+d)
Tot	35	72	107

Lo score Apgar viene calcolato per valutare lo stato di salute di un neonato. I neonati con punteggio alla nascita inferiore a 4 sono gravemente depressi e necessitano di intervento medico immediato, quelli con punteggio fra 4 e 6 sono moderatamente "a rischio", bisognosi di assistenza, vigilanza e ripetizione del test ogni 5 minuti, i neonati con punteggio fra il 7 e il 10 sono considerati normali.

RR=[2/16]/[33/91]=0.345

Il rischio relativo di avere un basso score nel gruppo che presentava crescita simmetrica all'ecografia **è il 35%** rispetto al gruppo che presentava una crescita asimmetrica all'ecografia.

Per calcolare l'intervallo di confidenza di RR bisogna passare alla scala logaritmica:

$$\log(\widetilde{RR}) \sim Normal(\log(RR), var(\log(RR)))$$

$$se(\log(\widetilde{RR})) = \sqrt{\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}}$$

$$se(\log(\widetilde{RR})) = \sqrt{\frac{1}{2} - \frac{1}{16} + \frac{1}{33} - \frac{1}{91}} = 0.676$$

$$\log(RR) = \log(0.345) = -1.065$$

$$\log(RR) \pm 1.96 * se(\log(RR)) = -1.065 \pm 1.96 * 0.676$$

$\log(RR)\ 95\%\ CI : [-2.389; 0.259]$ RR $95\%\ CI : [\exp(-2.389); \exp(0.259)]$=**[0.091; 1.296]**

Notare che il 95% CI contiene 1…

***Test di ipotesi su RR:** Testare l'ipotesi nulla H_0 : RR= 1 è equivalente ad effettuare un test di indipendenza in una tabella 2 × 2 (test Chi-quadrato). Nel nostro esempio, il Chi-quadrato ha un p value=0.08, quindi non c'è una evidenza sufficiente per rigettare l'ipotesi nulla.

Misure di Rischio: calcolo dell'intervallo di confidenza per OR

	Erosione dentale	Non erosione dentale	Tot
6+ ore di nuoto	32	118	150
<6 ore di nuoto	17	127	144
Tot	49	245	294

Studio caso-controllo in un campione di nuotatori per vedere se c'è una associazione tra l'esposizione al cloro e l'erosione dello smalto dentale.

OR=[32*127]/[118*17]=2.026

L'esposizione a 6+ ore di nuoto a settimana **è circa due volte maggiore** nel gruppo con erosione dello smalto dentale rispetto ai nuotatori senza erosione.

Per calcolare l'intervallo di confidenza di OR bisogna passare alla scala logaritmica:

$$\log(\widetilde{OR}) \sim Normal(\log(OR), var(\log(OR)))$$

$$se(\log(\widetilde{OR})) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

$$se(\log(\widetilde{OR})) = \sqrt{\frac{1}{32} + \frac{1}{118} + \frac{1}{17} + \frac{1}{127}} = 0.326$$

$$\log(OR) = \log(2.026) = 0.706$$

$$\log(OR) \pm 1.96 * se(\log(OR)) = 0.706 \pm 1.96 * 0.326$$

$\log(OR)$ 95% CI : $[0.067; 1.345]$ OR 95% CI : [exp(0.067); exp(1.345)]=**[1.069; 3.840]**

Notare che il 95% CI NON contiene 1…

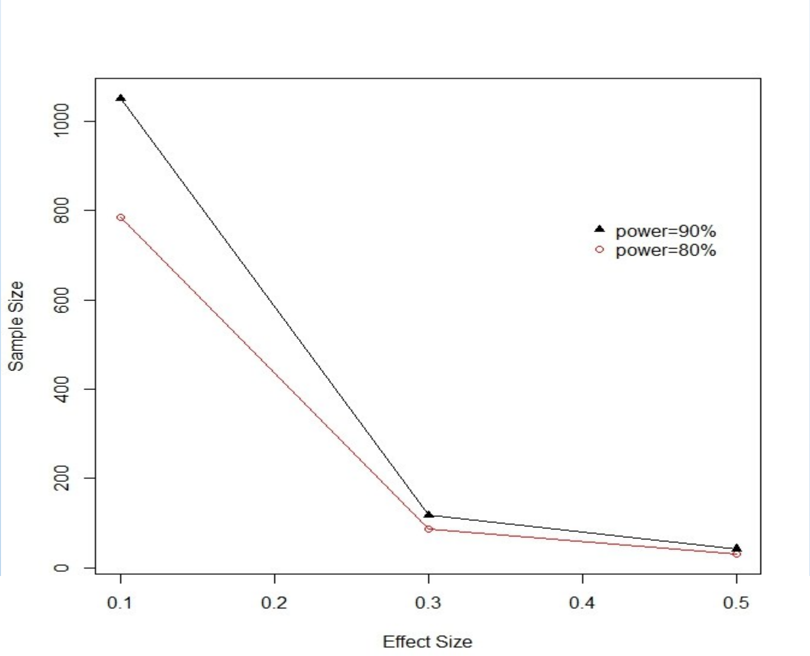
***Test di ipotesi su RR:** Testare l'ipotesi nulla H_0 : OR= 1 è equivalente ad effettuare un test di indipendenza in una tabella 2 × 2 (test Chi-quadrato). Nel nostro esempio, il Chi-quadrato ha un p value=0.03, quindi rigettiamo l'ipotesi nulla di indipendenza.

Sample size per stimare misure di rischio

	MALATI	**SANI**	**Totale**
ESPOSTI	a	b	a+b
NON ESPOSTI	c	d	c+d
Totale	a+c	b+d	N

In questo caso **l'effect size** corrisponde alla «forza» dell'associazione che ci attendiamo

(collegata alla dimensione dell'**odds ratio/relative risk**).



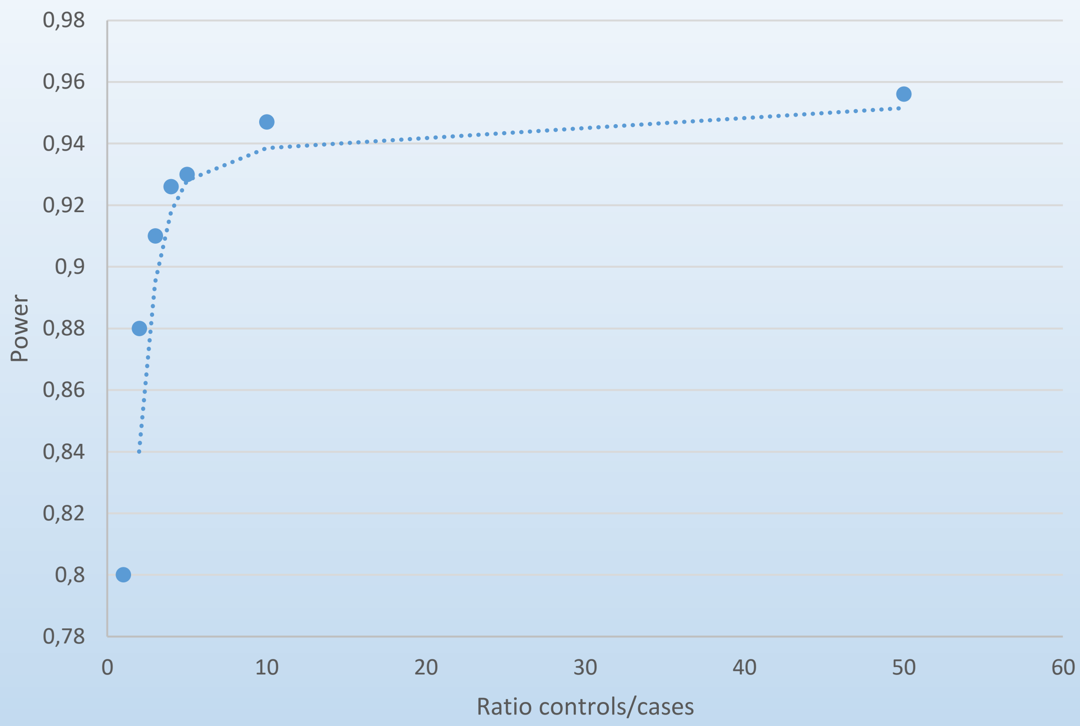
Size of effect	$w = \phi$	odds ratio*
small	.1	1.49
medium	.3	3.45
large	.5	9

Ingredienti «aggiuntivi»:

COORTE:
- Incidenza attesa di evento nel gruppo dei non esposti
- Rapporto non esposti/esposti

CASO/CONTROLLO:
- Prevalenza attesa di esposizione nel gruppo dei controlli
- Rapporto casi/controlli

CASI/CONTROLLI (oppure -> Esposti/Non Esposti):
raramente è necessario includere
più di 3 o 4 controlli (o non esposti) rispetto ai casi (o esposti).



Es: individuare un RR (Rischio Relativo) di almeno 2 per un outcome che avviene nel gruppo dei non esposti con una incidenza pari a 0.01, ad un livello di significatività del 5% ed una potenza dell'80%:
2319 esposti vs **2319** non esposti.

Ratio	Power
1:1	0,80
2:1	0,88
3:1	0,91
4:1	0,926
5:1	0,93
10:1	0,947
50:1	0,956

http://epitools.ausvet.com.au/content.php?page=cohortSS

ausvet — EpiTools epidemiological calcula

Home | Prevalence | Freedom | Studies | Diagnostics | Sampling | English

Sample size for a cohort study

Input Values

This utility calculates the sample size required for a cohort study, with specified levels of confidence and power and cohorts of equal size.

Inputs are the expected incidence in the unexposed cohort, the assumed relative risk, and the desired level of confidence and power for the detection of a significant difference between the two cohorts.

The program outputs the sample size required for the specified inputs. Also output is a summary table of sample sizes for a range of assumed incidence values and relative risks.

Expected incidence in unexposed: 0.01
Assumed relative risk: 2
Confidence level: 0.95
Power: 0.8
Submit

Sample size for a Cohort study

Analysed: Wed Apr 05, 2017 @ 01:04

Results

Sample size for specified values

Expected incidence in unexposed	0.01
Assumed relative risk	2
Confidence level	0.95
Power	0.8
Study type	Cohort study
Sample size per group	2316
Total sample size (both groups):	4632

Sample size (per group) for range of incidence values and relative risks

	P = 0.005	P = 0.01	P = 0.02	P = 0.03	P = 0.04	P = 0.05	P = 0.1
RR = 2	4671	2316	1139	746	550	432	197
RR = 3	1551	766	373	243	177	138	59
RR = 4	858	422	204	131	95	73	29
RR = 5	576	282	135	86	61	47	17
RR = 10	204	97	44	26	17	12	1

Studio di coorte

http://epitools.ausvet.com.au/content.php?page=case-controlSS

ausvet — EpiTools epidemiological calcula

Home | Prevalence | Freedom | Studies | Diagnostics | Sampling | English

Sample size for a case-control study

Input Values

This utility calculates the sample size required for a case-control study, with specified levels of confidence and power and case and control groups of equal size.

Inputs are the expected proportion exposed in the controls, the assumed odds ratio, and the desired level of confidence and power for the detection of a significant difference between the two groups.

The program outputs the sample size required for the specified inputs. Also output is a summary table of sample sizes for a range of assumed proportions exposed and odds ratios.

Expected proportion exposed in controls: 0.05
Assumed odds ratio: 2
Confidence level: 0.95
Power: 0.8
Submit

Sample size for a Case-control study

Analysed: Wed Apr 05, 2017 @ 01:07

Results

Sample size for specified values

Expected proportion in controls	0.05
Assumed odds ratio	2
Confidence level	0.95
Power	0.8
Study type	Case-control study
Sample size per group	513
Total sample size (both groups):	1026

Sample size (per group) for range of proportions exposed and odds ratios

	P = 0.005	P = 0.01	P = 0.02	P = 0.03	P = 0.04	P = 0.05	P = 0.1
OR = 2	4749	2395	1218	826	630	513	280
OR = 3	1586	801	409	279	213	174	97
OR = 4	882	446	228	156	120	98	56
OR = 5	596	302	155	106	82	67	38
OR = 10	217	110	57	39	30	25	15

Studio Caso-Controllo

Relative vs Absolute Effects…

The Telegraph

HOME | NEWS | SPOR

News

UK | World | Politics | Science | Entertainment | Pictures | Investigations | Brexit

News

Why binge watching your TV box-sets could kill you

share

Hours of inactivity slumped in front of a television **sharply raises the risk** of **dying from a blood clot in the lungs**, say scientists…

https://www.ahajournals.org/doi/10.1161/circulationaha.116.023671

RESEARCH ARTICLE | Originally Published 26 July 2016 | Check for updates

Watching Television and Risk of Mortality From Pulmonary Embolism Among Japanese Men and Women: The JACC Study (Japan Collaborative Cohort)

Toru Shirakawa, MD, Hiroyasu Iso, MD, PhD, MPH, Kazumasa Yamagishi, MD, PhD, Hiroshi Yatsuya, MD, PhD, Naohito Tanabe, MD, PhD, Satoyo Ikehara, PhD, Shigekazu Ukawa, PhD, and Akiko Tamakoshi, MD, PhD | AUTHOR INFO & AFFILIATIONS

Circulation • Volume 134, Number 4 • https://doi.org/10.1161/CIRCULATIONAHA.116.023671

8.2 cases each 100.000 p-years
vs
2.8 cases each 100.000 p-years

= 5.4 *extra* cases each 100.000 p-years …

Table. HRs for Mortality From Pulmonary Embolism According to Hours Spent Watching Television

	Time Spent Watching Television, h/d			
	<2.5	2.5–4.9	≥5.0	Increment by 2 h (95% CI), *P* Value
Cases, n/person-y	19/678 199	27/562 449	13/157 922	
Mortality rate per 100 000 person-y	2.8	4.8	8.2	
Age- and sex-adjusted HR	1.0	1.6 (0.9–2.8)	2.4 (1.2–4.9)	1.3 (1.0–1.8), 0.06
Multivariable HR*	1.0	1.7 (0.9–3.0)	2.5 (1.2–5.3)	1.4 (1.0–1.8), 0.04

CI indicates confidence interval; and HR, hazard ratio.

*Adjusted for age, sex, body mass index, history of hypertension, history of diabetes mellitus, smoking status, perceived mental stress, educational level, walking activity, and sports activity.

you can also say that
(if this effect is ***causal…***) *a subject*
would have to watch 5+ hours a day
for nearly 19.000 years to expect one
event to happen …

Excess risk

To convey an **absolute** measure of the impact of exposure on risk, the Excess Risk, denoted by ER, could be estimated:

$$ER = P(D|E) - P(D|\bar{E})$$

The Excess Risk uses the same basic components as the Relative Risk (and the Odds Ratio), but looks at the **absolute**, rather than relative, difference in risk levels.

The Relative Risk for lung cancer associated with cigarette smoking is about **6** times as great as the Relative Risk for CHD due to smoking.

On the other hand, the Excess Risk for CHD **is larger** since it is the **most common** disease.

Therefore, from a health policy or public health point of view, cigarette intervention programs may be more important in terms of their **impact** on CHD.

$$ER \begin{cases} <0 & \text{a greater risk for D when unexposed than when exposed} \\ =0 & \text{independence of D and E} \\ >0 & \text{a greater risk for D when exposed than when unexposed} \end{cases}$$

$$-1 \leq \mathrm{ER} \leq 1$$

Excess Risk* is the basis of an ***additive*** *model* for risk:

$$Risk_{Exposed} = Risk_{unexposed} + ER$$

Interpretation of the Excess Risk ??

*Note that in epidemiological studies ER is computed as a difference between *cumulative incidence* or *incidence rates* and expressed in terms of *#events* in a *population* over a *period of time*

$P(D) = P(D|\bar{E})$ All **not exposed**: number of cases $\# \, cases = \boldsymbol{N * P(D|\bar{E})}$

$\# \, cases = \boldsymbol{N * P(D)}$

$P(D) = P(D|E)$ All **exposed**: number of cases $\# \, cases = \boldsymbol{N * P(D|E)}$

Excess Risk = the "excess" number of cases when population members ***are all exposed*** as compared to them ***all being unexposed***.

Example: association between appendectomy and infections

Cumulative Incidence (CI) with appendectomy = 5.3% = 53/1000

Cumulative Incidence (CI) without appendectomy = 1.3% = 13/1000

Excess Risk (ER) = 40/1000= 4/100 (0.04%)

Interpretation:
Subjects who had appendectomy had 4 ***additional cases*** of infection per 100 people compared to subjects who did not have appendectomy.

There were 4 *excess infections* per 100 subjects in the group that had appendectomies, compared to the group without appendectomy.

Absolute effect

Excess Risk for infant mortality in the U.S. in 1991 associated with the mother's marital status:

	Mother's Marital Status	
Infant Mortality	Unmarried	Married
Death	16,712	18,784
Live at 1 year	1,197,142	2,878,421
Total	1,213,854	2,897,205

$$ER = \frac{16712}{1213854} - \frac{18784}{2897205} = 0.0073$$

Excess Risk for infant mortality associated with low birthweight:

	Birthweight		
Infant Mortality	Low Birthweight	Normal Birthweight	Total
Death	21,054	14,442	35,496
Live at 1 year	271,269	3,804,294	4,075,563
Total	292,323	3,818,736	4,111,059

$$ER = \frac{21054}{292323} - \frac{14442}{3818736} = 0.0682$$

Low birthweight is *more influential* than marital status on both the absolute and relative comparative scales.

We would expect the infant mortality **to increase by 7%** if all births exhibited low birthweight as compared to all those being of normal birthweight (w.r.t **0.7%** in case of marital status).

So…relative or absolute risk measures ? Better both!

Relative measures lose information on risk levels, so you can find relative risks *relatively low* associated with *very high absolute* differences, and viceversa.

Relative and absolute risk measures between incidence rates (per 100.000 pyrs) of disease in smokers and non-smokers:

	Smokers	Not Smokers	RR	ER*
Lung cancer	48.33	4.49	**10.8**	43.84
Cardiovascular disease	294.67	169.54	1.7	**125.13**

For this reason, it is important to estimate (especially in public health studies) also the absolute differences between risks / rates (note that this is possible *in some types of studies* but not in others).

*ER here is expressed as a difference between incidence rates: an «excess» of cases each 100.000 pyears

Decision Making:

For making informed clinical decisions or health policy recommendations for a **specific population**, the **absolute risk difference** (or measures derived from it, like the **Number Needed to Treat (NNT)**) is often considered more relevant.

Number Needed to Treat (NNT)

In uno studio clinico con una risposta binaria, come vivo o morto, ci sono molti modi per quantificare la differenza tra due trattamenti.

Ad esempio, è possibile utilizzare la differenza tra due proporzioni: $p_{new} - p_{old}$

p_{new} = proporzione di successi con il nuovo trattamento ; p_{old}= proporzione di successi con il vecchio trattamento

$$NNT = \frac{1}{p_{new} - p_{old}}$$

NNT : # di pazienti da sottoporre al nuovo trattamento per ottenere **un ulteriore successo** rispetto al vecchio trattamento.

$$1 \leq NNT < \infty \begin{cases} \dfrac{1}{1-0} & \text{Il nuovo trattamento } \textbf{\textit{è sempre efficace}} \text{ mentre il vecchio non lo è mai} \\[2ex] \dfrac{1}{0} & \text{Il nuovo trattamento } \textbf{\textit{non è in alcun modo efficace}} \text{ (la differenza è pari a zero)} \end{cases}$$

OPZIONALE

Potremmo avere la situazione opposta, il nuovo trattamento è dannoso, e quindi ottenere un valore negativo:

NNT < 0 → NNH= ***number needed to harm*** (numero necessario per avere un danno)

Esempio:
In uno studio sulla streptomicina, nel quale i soggetti erano selezionati tra i pazienti affetti da tubercolosi polmonare, la proporzione di sopravviventi a 6 mesi era 93% nel gruppo trattato vs 73% nei controlli.

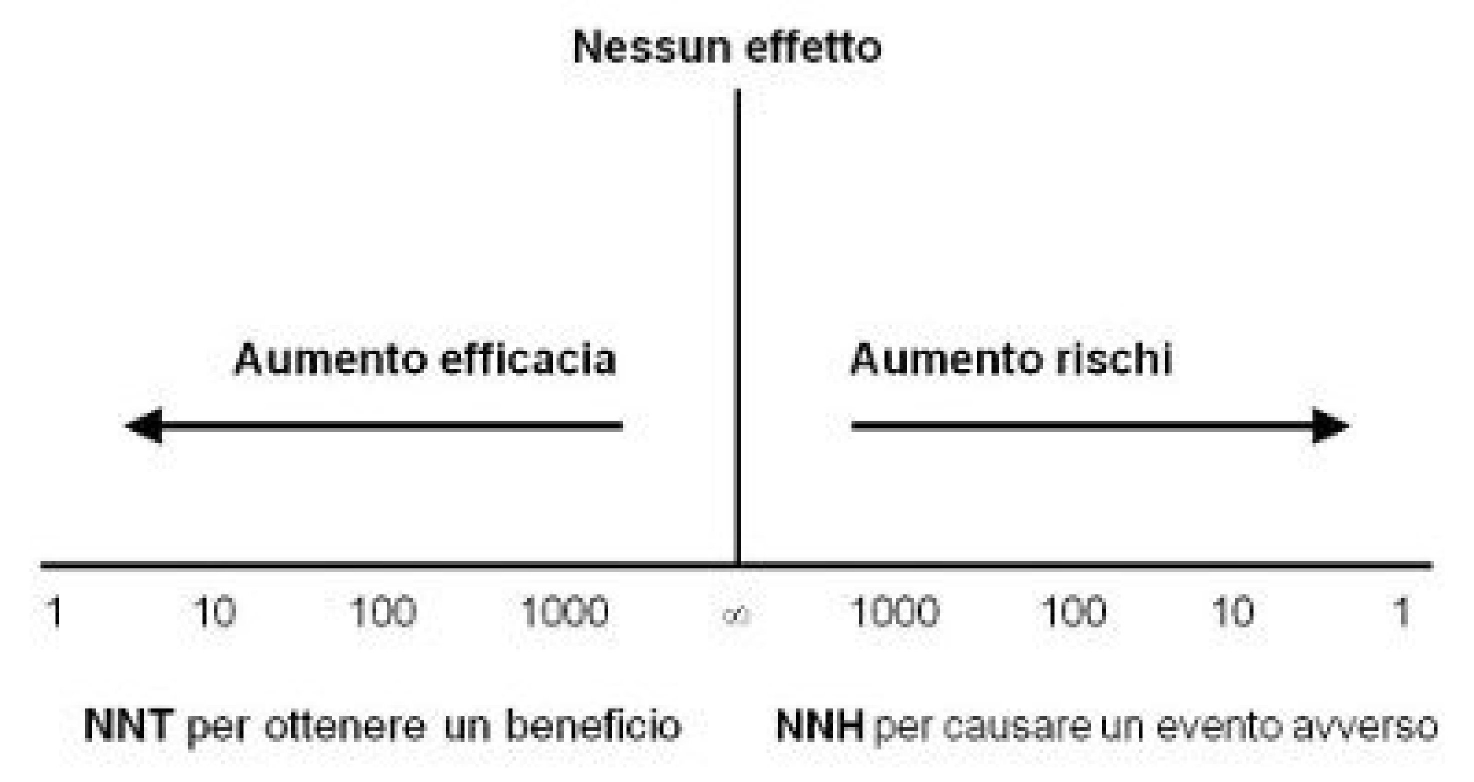
$$NNT = \frac{1}{p_{new} - p_{old}} = \frac{1}{0.93 - 0.73} = \frac{1}{0.20} = 5$$

Quindi il numero necessario di pazienti da trattare per prevenire un decesso a 6 mesi era pari a 5.

E’ possibile calcolare anche gli intervalli di confidenza attorno a NNT, ma non entriamo in questo dettaglio*.

*Sorgono dei problemi qualora la differenza tra le proporzioni non sia statisticamente significativa, perché l’intervallo di confidenza in quel caso contiene lo zero, in tal caso infinito è un possibile valore per NNT…così come sono ammissibili valori negativi.

OPZIONALE



Al **diminuire** del NNT aumenta l'efficacia del trattamento, per cui 1 è il NNT ideale: un successo terapeutico per ciascun paziente trattato (misura di **efficacia**).

All'**aumentare** del NNH si riduce la probabilità di eventi avversi e aumenta la sicurezza del trattamento, per cui il NNH ideale tende all'infinito, documentando l'assenza di eventi avversi (misura di **sicurezza**).