

# GLOSSARIO EBM

<http://aifa.clinev.it/metodologia.php?let=V>

## Analisi di regressione

➤ L'analisi di regressione è una tecnica statistica che studia l'eventuale presenza di una relazione tra una variabile definita a priori (variabile dipendente  $y$ ) e una variabile indipendente o predittiva  $x$  (regressione lineare semplice); se le variabili predittive sono più di una si parla di regressione lineare multipla (tecnica multivariata). Il risultato dell'analisi di regressione esprime la dipendenza funzionale tra le variabili, sotto forma di relazioni (per esempio relazione lineare, quadratica, cubica). In pratica, nel caso di una relazione lineare al crescere della variabile indipendente cresce anche il valore di quella dipendente di un fattore 1, 2, ... 10 eccetera; la pendenza della curva (coefficiente angolare) indica quanto è grande la dipendenza tra le 2 variabili (una maggiore pendenza rappresenta una più forte relazione). Il fattore è espresso al quadrato o al cubo se la relazione è quadratica o cubica.

## Analisi limitata ai soggetti rimasti nello studio

➤ Quando si conduce uno studio, l'analisi dei dati può essere eseguita su tutti i soggetti originariamente inclusi (analisi secondo il principio dell'intenzione al trattamento). Altrimenti, l'analisi può essere limitata solo ai soggetti per i quali si dispone di osservazioni complete fino alla conclusione del follow up. Questo secondo tipo di analisi, che potrebbe apparire più logica dal punto di vista clinico in quanto permette di basare le conclusioni su dati certi, è tuttavia suscettibile di diversi tipi di errore. Infatti i soggetti che sono stati persi al follow up potrebbero aver abbandonato lo studio per motivi collegati all'esito o alla gravità della malattia. Se così fosse, basando l'analisi solo sui soggetti rimasti in follow up si introdurrebbe nello studio un errore sistematico o *bias*, in quanto questi soggetti sarebbero sistematicamente diversi da quelli che invece si sono persi. Si tratta quindi di un tipo di analisi che dovrebbe essere evitata a favore, quando possibile, dell'analisi secondo il principio dell'intenzione al trattamento. Un'ulteriore possibilità è rappresentata dall'analisi secondo il peggiore scenario, con la quale si ipotizza per esempio che tutti i persi al follow up tra i soggetti assegnati al trattamento abbiano avuto un esito negativo, mentre i persi al follow up tra gli appartenenti al gruppo di controllo abbiano avuto un esito positivo. Se nonostante tale assunzione sfavorevole l'effetto del trattamento persiste, si conclude per il beneficio del trattamento.

## Analisi secondo il principio dell'intenzione al trattamento

➤ Quando si realizza uno studio è importante che l'analisi dei dati venga effettuata: 1) su tutti i partecipanti inizialmente reclutati, indipendentemente dal fatto che essi siano realmente rimasti all'interno dello studio fino alla conclusione del follow up; 2) analizzando ogni soggetto secondo il gruppo

assegnato dalla randomizzazione<sup>7</sup>, indipendentemente dal fatto che abbia o non abbia assunto tale trattamento. Condizione per eseguire questo tipo di analisi è la conoscenza dell'esito finale di ogni soggetto. Particolare attenzione va rivolta a chi si è ritirato dallo studio (*dropout*) in quanto il comportamento di questi soggetti può essere diverso da quello di chi è invece rimasto nello studio fino alla fine per motivi legati alla malattia (gravità) o al trattamento (tollerabilità, tossicità eccetera). Un trattamento potrebbe, infatti, essere molto efficace ma produrre anche un elevato numero di *dropout* a causa degli effetti collaterali; la sua efficacia complessiva e la sua applicabilità ne uscirebbero quindi ridimensionate. L'analisi secondo il principio dell'intenzione al trattamento assume particolare importanza nei casi in cui alcuni pazienti abbiano ricevuto il trattamento previsto per l'altro gruppo (per esempio: pazienti assegnati per randomizzazione a una terapia antibiotica che non assumono l'antibiotico, pazienti randomizzati a non assumere la terapia che invece assumono il farmaco). Anche in questi casi si consiglia di analizzare i soggetti secondo il gruppo di randomizzazione (*intention to treat*) e non secondo la terapia effettivamente assunta (*analisi per protocollo*). L'analisi secondo il principio dell'intenzione al trattamento presenta alcuni vantaggi: si tratta di un'analisi conservativa, che tende ad attutire le differenze tra i gruppi; rispecchia più fedelmente quanto poi avverrà nella pratica quotidiana; permette di rispettare e di mantenere la ripartizione stabilita dalla randomizzazione.

### **Aumento del rischio assoluto (ARI)**

<sup>7</sup>In uno studio, l'aumento del rischio assoluto (o aumento assoluto del rischio) rappresenta la differenza nel rischio assoluto<sup>7</sup> fra gruppo sperimentale e gruppo di controllo<sup>7</sup>. Questa misura è impiegata quando il rischio nel gruppo sperimentale è superiore al rischio nel gruppo di controllo e viene calcolata sottraendo il rischio assoluto nel gruppo di controllo dal rischio assoluto nel gruppo sperimentale. Si tratta di un parametro che non dà informazioni sull'aumento proporzionale fra i 2 gruppi; per avere quest'ultima informazione bisogna calcolare l'aumento del rischio relativo<sup>7</sup>. Un valore superiore a 0 indica effettivamente un rischio, mentre un valore negativo ha il significato di un effetto benefico o protettivo; un valore pari a 0 indica l'indifferenza tra i 2 trattamenti (esposizioni).

### **Aumento del rischio relativo (RRI)**

<sup>7</sup>L'aumento del rischio relativo (o aumento relativo del rischio) è l'incremento proporzionale del rischio tra i soggetti che fanno parte del gruppo sperimentale rispetto a quello presente tra i soggetti del gruppo di controllo<sup>7</sup>. L'aumento del rischio assoluto<sup>7</sup> (differenza di rischi tra i 2 gruppi) viene rapportato al rischio del gruppo di controllo ed espresso in termini percentuali. Un valore superiore a 0 indica un aumento di rischio, mentre un valore negativo ha il significato di un effetto benefico o protettivo; un valore pari a 0 indica l'indifferenza tra i 2 trattamenti (esposizioni).

## Bias di pubblicazione

Il *bias* di pubblicazione è una distorsione esistente nella letteratura medica dovuta al fatto che tendono a essere pubblicati più facilmente gli studi che hanno prodotto risultati positivi. In altre parole, gli studi i cui risultati mostrano l'efficacia di un trattamento hanno maggiori probabilità di essere accettati e pubblicati dalle riviste mediche rispetto agli studi i cui risultati mostrano che il trattamento è inefficace. A causa del *bias* di pubblicazione i dati presenti in letteratura tendono quindi a enfatizzare i risultati positivi; per questo motivo il *bias* di pubblicazione è una grande insidia per le revisioni della letteratura, se queste vengono basate solamente sugli studi pubblicati. Le revisioni sistematiche basate sia sugli studi pubblicati sia su quelli non pubblicati (ma reperibili sotto forma di *abstract* a congressi, tesi di laurea, dati contenuti nei documenti per la registrazione dei farmaci eccetera) dovrebbero invece risentire in misura minore del *bias* di pubblicazione. E' chiaro che la ricerca degli studi non pubblicati è complessa e difficilmente può essere esaustiva. Dagli studi che hanno valutato l'importanza del *bias* di pubblicazione non sono sempre emerse risposte univoche, per cui la sua influenza dovrebbe essere valutata caso per caso.

## Clinicamente rilevante

Con questa espressione ci si riferisce alla rilevanza e alla ricaduta clinica (per il singolo paziente e per gruppi di pazienti) dell'effetto di un determinato intervento. La rilevanza clinica di un intervento non deve essere confusa con la sua significatività statistica. Idealmente, quando si misura l'effetto di un intervento all'interno di uno studio si va alla ricerca di un effetto che sia al tempo stesso clinicamente rilevante e statisticamente significativo. Tuttavia rilevanza clinica e significatività statistica dipendono da fattori diversi. Piccole differenze, se misurate su campioni di pazienti molto grandi, possono essere statisticamente significative ma irrilevanti dal punto di vista clinico; per esempio, una differenza nei valori di pressione arteriosa di pochi mmHg in seguito a un trattamento può essere statisticamente significativa ma irrilevante dal punto di vista clinico a causa dell'esiguità dell'effetto. Al contrario, le differenze rilevate tra i gruppi studiati possono non essere statisticamente significative se i campioni sono molto piccoli ma indicare una differenza che, se non fosse dovuta al caso, potrebbe avere grande importanza clinica.

## Differenza media pesata

Misura della dimensione dell'effetto usata nelle metanalisi di studi nei quali gli esiti sono valutati con variabili quantitative (come i punteggi sintomatologici o la statura) e non come variabili qualitative (come la morte o l'infarto miocardico). Le differenze medie nell'esito fra i gruppi studiati sono ponderate per tener conto delle dimensioni campionarie differenti e della diversa precisione fra gli studi. E' un numero assoluto e dunque assume l'unità di misura dell'esito misurato originariamente.

## Differenza media standardizzata

➤ Misura della dimensione dell'effetto ➤ usata nelle metanalisi ➤ di studi nei quali gli esiti sono valutati con variabili quantitative (come i punteggi sintomatologici o la statura) e non come variabili qualitative (come la morte o l'infarto miocardico) e nei quali sono stati impiegati strumenti di valutazione diversi nel calcolare la dimensione dell'effetto. Per esempio, se sono state usate scale di valutazione del dolore che hanno sistemi di punteggio diversi (da 1 a 10, a 50 o a 100) nel computo della stima combinata bisogna tener conto di queste differenze: i punteggi devono essere standardizzati in un'unica scala che riporti tutte le misure in deviazioni standard rispetto alle medie di ogni singolo elemento.

## Dimensione dell'effetto

➤ Molti metodi vengono usati per misurare la dimensione di un effetto (*effect size*), sia nei singoli studi sia nelle metanalisi ➤. Ne sono esempi il rischio relativo ➤ e l'odds ratio ➤ per esiti dicotomici; la differenza media standardizzata ➤ o la differenza media pesata ➤ per variabili quantitative (come i punteggi sintomatologici o la statura).

## Errori e distorsioni negli studi clinici

Il complesso sistema per la misurazione della malattia, dell'esposizione, del trattamento, dell'esito, nonché la messa in atto di strategie di conduzione dello studio, a volte complesse e articolate, mira a un unico risultato: poter distinguere l'effetto del trattamento (o più in generale dell'intervento) rispetto a tutti quei fattori, detti confondenti, che potrebbero inserirsi nella catena degli eventi.

Il *bias* (traducibile come "distorsione") è un errore sistematico presente in uno studio; tale errore si ripercuote sui risultati, determinando uno scarto tra i risultati ottenuti e quelli che si sarebbero dovuti ottenere in assenza di *bias*. Il *bias* è frutto del confondimento (che si verifica quando è presente un fattore estraneo, di confondimento appunto, che è associato sia all'esposizione-trattamento sia all'esito).

Vi sono molte possibili classificazioni dei *bias*, ma le categorie più importanti sono:

- **bias di selezione**, che si verifica se il campione indagato è stato scelto e assemblato in modo errato, con differenze sistematiche nell'arruolamento (non comparabilità delle popolazioni);
- **performance bias**, dovuto a differenze sistematiche nella somministrazione degli interventi, trattamenti o esposizioni (non comparabilità degli effetti estranei);

- **bias di misurazione**, se i metodi di misurazione non sono ben tarati, o validi, e tale errore si verifica in misura sistematicamente differente nei diversi gruppi; per esempio, il *detection bias* è dovuto a differenze sistematiche nella valutazione degli esiti a causa di classificazione errata, differente attenzione nell'osservare eccetera. Può essere classificato in questa categoria anche il cosiddetto *attrition bias*, legato a perdite al follow up non casuali ma differenziate in modo sistematico nei diversi gruppi.

I *bias* vengono ridotti al minimo dagli studi randomizzati e in doppio cieco grazie alla randomizzazione, alla cecità, all'analisi secondo il principio dell'intenzione al trattamento. Negli studi non randomizzati diventa cruciale il controllo del cosiddetto confondimento, con tecniche come l'appaiamento o *matching* (a ogni soggetto scelto si affianca un altro simile per caratteristiche quali sesso, età, condizione sociale), la restrizione (si includono solo certe sottoclassi, per esempio i maschi, i fumatori, gli anziani), la stratificazione (per esempio, tenendo fissa l'età si studia l'effetto del sesso o di altre variabili). In fase di analisi, la tecnica più usata è l'analisi multivariata. Si tratta di una metodologia statistica che valuta l'effetto di un insieme di variabili indipendenti (o predittive) su una (o più) variabili dipendenti (o di esito). Tra le tecniche più note, la regressione multipla lineare (per esiti misurabili lungo una scala numerica), la regressione logistica multipla (per esiti espressi in forma binaria sì/no), l'analisi discriminante, l'analisi dei cluster, l'analisi delle corrispondenze. Il principale vantaggio dell'analisi multivariata consiste nel poter apprezzare, simultaneamente e indipendentemente, gli effetti di ognuna delle variabili considerate, depurati delle reciproche interferenze. Inoltre è possibile valutare i diversi effetti delle variabili in grado di influenzarsi (interazione) positivamente o negativamente. L'analisi multivariata è pertanto una delle tecniche più robuste di controllo a posteriori del confondimento, del quale è in grado di tener conto: gli effetti stimati sono "aggiustati", considerano cioè la presenza contemporanea di tutte le altre variabili prese in esame. Generalmente, per ogni variabile inclusa in un modello di analisi multivariata vengono presentate la significatività dell'effetto, espressa dal valore di un test statistico e dal relativo valore di P e la forza o intensità dell'effetto, attraverso il valore dei cosiddetti coefficienti beta.

## Eterogeneità

🚩 In una metanalisi 🚩 l'eterogeneità indica l'esistenza di differenze tra gli studi. Può essere dovuta all'utilizzo di metodi statistici differenti nei vari studi (eterogeneità statistica) o al fatto che le valutazioni sono state effettuate su differenti tipi di pazienti, di trattamenti o di esiti (eterogeneità clinica). L'eterogeneità può rendere inaffidabile o inappropriata la combinazione quantitativa dei dati prodotti dai vari studi. Quando si sospetta l'esistenza di eterogeneità tra gli studi inclusi in una metanalisi (sospetto che può ricevere conferma da appositi test statistici) le ragioni di tale eterogeneità andrebbero sempre indagate. In caso di eterogeneità è consigliabile per l'analisi statistica l'utilizzo di modelli a effetti variabili.

## Evento

➤ Con il termine evento ci si riferisce al verificarsi in uno studio di un esito dicotomico come vivo/morto, migliorato/peggiolato eccetera.

## Falso negativo

➤ Un falso negativo è un risultato errato di un test diagnostico, tale per cui un soggetto affetto dalla condizione/malattia indagata ha un risultato negativo del test diagnostico effettuato per rilevarne la presenza.

## Falso positivo

➤ Un falso positivo è un risultato errato di un test diagnostico, tale per cui un soggetto non affetto dalla condizione/malattia indagata ha un risultato positivo del test diagnostico effettuato per rilevarne la presenza.

## Gruppo di controllo (o di confronto)

➤ In generale, è il gruppo con cui si confronta un gruppo di interesse in uno studio: negli studi prospettici non randomizzati il gruppo di controllo è costituito da soggetti non esposti; negli studi caso-controllo ➤ da soggetti non affetti dalla condizione studiata; negli studi sperimentali ➤, da pazienti che non ricevono il trattamento valutato. I 2 (o più) gruppi studiati devono avere caratteristiche simili che li rendano confrontabili. Nel caso di uno studio randomizzato ➤, i pazienti del gruppo di controllo sono assegnati a ricevere: il trattamento ritenuto lo standard di riferimento per quella malattia (in questo caso si parla di gruppo di controllo che riceve la terapia standard); oppure, se non esistono terapie efficaci, un placebo ➤ o nessun trattamento (in questo caso si parla di gruppo di controllo non trattato).

## Hazard ratio (HR)

➤ L'hazard ratio esprime il rapporto tra rischi nelle analisi di sopravvivenza. In pratica esprime, in un periodo di follow up definito, il rapporto tra i rischi di un evento (morbilità o mortalità) in un gruppo di soggetti trattati rispetto a un altro di non trattati. Viene spesso utilizzato nelle metanalisi ➤, nelle quali confronta la sopravvivenza globale (dal tempo zero fino a un valore di tempo prefissato) tra 2 gruppi di trattamento esaminati in diversi studi clinici.

## Incidenza

➤ L'incidenza è la frequenza di comparsa, in un periodo di tempo definito, di nuovi casi all'interno di una popolazione prestabilita. Tutti e tre gli elementi sono importanti: il numero dei casi, il numero delle persone che compongono quella popolazione e il periodo di tempo preso in esame. Il denominatore espresso dal numero di nuovi casi nel periodo considerato viene infatti rapportato al numero di soggetti presenti all'inizio del periodo (incidenza cumulativa o rischio) oppure al cumulo di "tempo per persone"

complessivamente esposto al rischio nel periodo studiato (tasso di incidenza in senso stretto).

## **Intervallo (limiti) di confidenza**

➤ L'intervallo di confidenza è un intervallo di valori, calcolato dai dati del campione, che con una certa probabilità (pari al livello di confidenza, per esempio 95%, 99%) permette di essere nel giusto affermando che il vero valore del parametro studiato è compreso all'interno dell'intervallo stesso. Bisogna ricordare che uno studio non analizza mai la popolazione intera, ma sempre un campione. I risultati ottenuti non rappresentano pertanto il "valore vero" di quel parametro nella popolazione generale, ma un valore stimato attraverso i risultati dello studio. Un intervallo di confidenza al 95% i cui limiti cadano dalla stessa parte rispetto al valore di indifferenza (0 per le differenze, 1 per i rapporti) permette di concludere per una significatività dell'effetto del trattamento; ha lo stesso significato di un valore di  $P$  inferiore a 0,05 ma è più informativo del test di significatività statistica. A differenza infatti di quest'ultimo, che dà un risultato soglia (statisticamente significativo sì/no), i limiti dell'intervallo di confidenza forniscono una misura quantitativa dell'entità clinica dell'effetto.

## **Metanalisi**

➤ E' la combinazione quantitativa dei dati provenienti da diversi studi cui si può arrivare nell'ambito di una revisione sistematica ➤ utilizzando specifiche tecniche statistiche. Questa combinazione si realizza non come somma semplice dei risultati dei singoli studi come se fossero un unico esperimento, ma assegnando alle stime dell'effetto ottenute in ogni singolo studio un peso proporzionale sia alla numerosità del campione (gli studi più grandi hanno un peso maggiore degli studi più piccoli), sia alla frequenza degli eventi osservati (peso maggiore agli studi con maggior numero di eventi). I termini metanalisi e revisione sistematica vengono erroneamente usati come sinonimi; molte revisioni sistematiche non procedono all'esecuzione di una metanalisi in quanto la diversità dei disegni di studio o la natura dei dati non lo consente. Quindi, nei capitoli di CE, il lettore troverà quasi sempre il riferimento a revisioni sistematiche mentre più raramente sono disponibili vere e proprie metanalisi quantitative. Le metanalisi, se ben condotte, possono fornire informazioni che vanno al di là dei risultati raggiunti con i singoli studi. Combinando i risultati di numerosi studi si aumenta infatti il numero dei soggetti disponibili e quindi la potenza statistica dei confronti. Inoltre, una metanalisi di studi condotti su quesiti relativamente simili ma su pazienti o *setting* assistenziali diversi permette di saggiare la generalizzabilità e la coerenza dei risultati.

## **Misura degli esiti**

L'esito di un intervento in un gruppo viene solitamente espresso come proporzione di soggetti con esiti favorevoli o sfavorevoli rispetto al totale dei soggetti considerati (variabili qualitative) o come media di un certo parametro

di interesse tra i soggetti del gruppo (variabili quantitative). **Il confronto tra 2 gruppi ha come misura il rapporto o la differenza tra proporzioni (variabili qualitative) o la differenza tra medie (variabili quantitative).** **Rischio** (sinonimo di **probabilità**) e **odds** sono entrambi **misure di frequenza**: la frequenza con cui si presenta un fenomeno, un evento, una caratteristica (per esempio la malattia X, il fattore di rischio Y eccetera). Una frequenza può essere misurata in un certo istante, come negli studi trasversali, e in questo caso si parla di prevalenza (per esempio, quanti soggetti in un certo momento sono portatori di calcolosi biliare). E' molto importante anche conoscere la frequenza con cui gli eventi si accumulano nel tempo. Negli studi longitudinali ci si domanda quanti dei soggetti facenti parte di un gruppo o campione inizialmente sano si ammalano durante un periodo (per esempio, quanti soggetti maschi di 40 anni si ammalano di calcolosi biliare entro 5 anni). Questa misura si chiama anche rischio e corrisponde esattamente alla probabilità; non va confusa con l'incidenza, che invece rapporta i nuovi casi di malattia alla massa di "tempo per persone" esposte al rischio. Per esempio, in uno studio prospettico l'incidenza di cancro del polmone nei fumatori potrebbe essere pari a 91 casi ogni 100.000 anni persona, intendendo con questa espressione che 91 casi di cancro del polmone si generano da 100.000 fumatori in un anno, o da 10.000 fumatori in 10 anni. Supponiamo che in una divisione chirurgica, su 40 soggetti presenti, 12 siano portatori di calcolosi biliare. In questo esempio di rilevazione trasversale, il valore della probabilità di selezionare un caso di calcolosi attraverso un'estrazione casuale dalla lista dei pazienti ricoverati, è dato dal rapporto di prevalenza  $12/40=0,3=30\%$ . Il numeratore, 12, è una parte del denominatore, 40. Il valore minimo della probabilità è 0 nel caso in cui nessuno dei soggetti considerati abbia la caratteristica in questione ( $P=0/40=0=0\%$ ); il massimo valore possibile è invece 1, nel caso in cui tutti i soggetti abbiano la caratteristica studiata ( $P=40/40=1=100\%$ ). L'odds invece è il rapporto tra il numero di eventi e il numero di non eventi; nel caso considerato,  $12/28=0,42857$ . Identico risultato si ha se si considera il rapporto tra la probabilità dell'evento e la probabilità del non evento; nel caso considerato,  $(12/40)/(28/40)=0,42857$ . L'odds esprime un contrasto tra chi possiede una caratteristica e chi non la possiede, tra portatori e non portatori di calcolosi biliare; si calcola come un rapporto, ma tra 2 grandezze che si escludono a vicenda: se un soggetto si ammala, esce dal denominatore e va ad alimentare il numeratore; i valori dell'odds possono dunque spaziare da 0 fino a valori molto alti, tendenti all'infinito. Occorre ricordare che per frequenze molto basse (per esempio una malattia rara) i valori di probabilità e odds praticamente coincidono, mentre all'aumentare delle frequenze i 2 valori divergono notevolmente.

**Rapporti tra probabilità e rapporti tra odds:** Il rapporto tra 2 probabilità è detto anche rapporto di prevalenza nel caso di studi trasversali o rischio relativo nel caso di studi prospettici. Sia il rischio relativo sia l'odds ratio esprimono la forza dell'associazione tra l'esposizione e la malattia. In un certo senso sono dunque sinonimi. L'interpretazione di questi indici è semplice: un valore superiore a 1 indica un aumento di rischio legato all'esposizione; un



valore pari a 1 indica la non influenza dell'esposizione; un valore inferiore a 1 indica una riduzione di rischio legato all'esposizione.

**Differenze tra rischi:** L'efficacia del trattamento in studio può essere valutata rispetto al trattamento di controllo come aumento o riduzione di frequenza dell'esito considerato. Tale differenza tra rischi può essere valutata sia in termini assoluti (aumento o riduzione del rischio assoluto, o anche aumento assoluto o riduzione assoluta del rischio) sia in termini relativi, rapportando la differenza assoluta al valore di rischio del gruppo di controllo (aumento o riduzione del rischio relativo, o anche aumento relativo o riduzione relativa del rischio). Nel caso di una differenza in termini assoluti, si può calcolare come reciproco ( $1/\text{differenza assoluta}$ ) il numero di soggetti che devono essere trattati per ottenere un successo in più nei confronti del trattamento di controllo (numero di casi da trattare) o il numero di soggetti che devono essere trattati per osservare un effetto avverso in più nei confronti del trattamento di controllo.

**Differenze tra medie:** Nel caso di confronti tra variabili quantitative, il confronto avviene solitamente tra le medie dei 2 gruppi corredato dal valore della deviazione standard (per esempio: FEV1 medio nei trattati 1,62 l, deviazione standard 12,4 l; FEV1 medio nei non trattati 1,42 l, deviazione standard 15,3 l: differenza media 0,2 l, deviazione standard 0,28).

Misure dei test diagnostici

Nell'ambito della misurazione della performance dei test diagnostici si utilizzano le seguenti definizioni:

- sensibilità del test: è la probabilità che data la presenza di una malattia, il test per rilevarla sia positivo;
- specificità del test: è la probabilità che, in assenza di una malattia M, il test per rilevarla sia negativo;
- potere predittivo positivo: è la probabilità di malattia dopo che si è eseguito il test nel caso in cui il test sia positivo;
- potere predittivo negativo: è la probabilità di malattia dopo che si è eseguito il test nel caso in cui il test sia negativo.

E' bene ricordare che:

- l'aumento della sensibilità del test tende a ridurre la frequenza dei falsi negativi, dunque ad aumentare il potere predittivo del test in caso di esito negativo; un test sensibile è utile per escludere, se negativo, la presenza della malattia;
- l'aumento della specificità del test tende a ridurre la frequenza dei falsi positivi, dunque ad aumentare il potere predittivo del test in caso di esito positivo; un test specifico è utile per confermare, se positivo, la presenza della malattia.

Si definisce rapporto di verosimiglianza il rapporto tra la frequenza con cui un certo valore del test si presenta tra i malati e la frequenza con cui lo stesso

valore si presenta nei sani. Per esempio, se osservassimo un valore di AST compreso tra 120 e 180 nel 60% dei casi di epatite cronica e nel 5% dei soggetti non affetti da epatite cronica, allora il rapporto di verosimiglianza sarebbe pari a  $60/5=12$ . In questo caso la probabilità di trovare questo livello di AST tra i malati è pari a 12 volte la probabilità di trovare lo stesso livello tra i sani.

## Misure di associazione

Le misure di associazione quantificano la relazione esistente tra un fenomeno (esito clinico, malattia, morte) e l'esposizione a un particolare fattore causale di malattia o a un trattamento. Ci sono 2 tipi principali di misure di associazione: 1) misure assolute, come la riduzione del rischio assoluto, e il suo reciproco, il numero di casi da trattare o NNT, che misurano la differenza tra esposti e non esposti nella frequenza di comparsa del fenomeno studiato; 2) misure relative, come la riduzione del rischio relativo, il rischio relativo e l'odds ratio, che misurano il rapporto tra esposti e non esposti nella frequenza di comparsa del fenomeno studiato. Le misure assolute indicano l'effetto assoluto di un intervento: per esempio, quanti casi in meno di una determinata malattia si possono ottenere eliminando un fattore eziologico. Le misure relative danno invece una misura della forza dell'associazione ma non permettono di valutare l'effetto assoluto di un intervento, che dipende – oltre che dalla forza dell'associazione – dalla frequenza di base di quella determinata malattia/condizione in una determinata popolazione.

## Modello a effetti casuali (*random effects*)

Modello di metanalisi che considera i lavori selezionati come un campione casuale di tutti i possibili studi sull'argomento. Vengono in questo caso inglobate nelle stime entrambe le componenti della variabilità, quella all'interno dei singoli studi e quella tra studi diversi (eterogeneità).

## Modello a effetti fissi (*fixed effects*)

Modello di metanalisi che considera i lavori selezionati come frutto di una ricerca esaustiva, cioè come la popolazione di tutti i lavori esistenti. L'unica fonte di variabilità è dunque la variabilità entro studi, mentre non viene presa in considerazione la variabilità tra studi (eterogeneità).

## Numero di casi da trattare (NNT)

L'acronimo NNT sta per *number needed to treat*. L'NNT è una misura dell'efficacia di un trattamento, e indica il numero di pazienti che devono ricevere uno specifico trattamento (per esempio un nuovo farmaco), anziché il trattamento di controllo, perché uno di loro ne tragga beneficio. Si consideri, per esempio, l'uso di un antibiotico nell'otite media acuta in età pediatrica, valutata in base alla persistenza del dolore tra il secondo e il settimo giorno di malattia, supponendo che la persistenza del dolore si verifichi nel 14% dei non

trattati e nel 9% dei trattati. La differenza del rischio assoluto è:  $0,14 - 0,09 = 0,05$ , con  $NNT\ 1/0,05 = 20$ . In altri termini, su 100 bambini trattati 86 sarebbero stati liberi dal dolore anche senza trattamento (remissione spontanea), 9 avrebbero ancora dolore nonostante l'assunzione dell'antibiotico, e 5 soltanto starebbero meglio grazie al trattamento ( $100:5 = 20$ ). Non è ovviamente possibile sapere in anticipo quali dei 100 bambini entreranno fra i 5 fortunati, bisogna dunque trattarli tutti e 100. L'NNT è un indice indiretto dell'efficacia dei trattamenti: più piccolo è il suo valore, maggiore è l'efficacia di quel trattamento e viceversa. Un NNT di 2 o 3 vorrebbe dire che basta trattare 2 o 3 pazienti perché 1 ne tragga beneficio, situazione assolutamente auspicabile ma purtroppo rara. L'NNT può essere ben rappresentato anche graficamente, come viene proposto nel sito Internet <http://www.nntonline.net/>.

## **Numero di casi da trattare per osservare un evento avverso (NNH)**

🚩 L'acronimo NNH sta per *number needed to harm*. L'NNH è una misura del danno che può essere indotto da un trattamento e indica il numero di pazienti che devono ricevere uno specifico trattamento (per esempio un nuovo farmaco), anziché il trattamento di controllo, perché uno di loro abbia un effetto avverso. Si consideri, per esempio, l'uso di un antibiotico nell'otite media acuta in età pediatrica, supponendo che si abbiano effetti avversi (vomito, diarrea o la comparsa di un rash cutaneo) nel 9% dei trattati e nell'1% dei non trattati. La differenza del rischio assoluto è:  $0,09 - 0,01 = 0,08$ , con  $NNH\ 1/0,08 = 12$ . In altri termini, su 100 bambini trattati 8 pazienti in più manifesteranno vomito, diarrea o la comparsa di un rash cutaneo ( $100:8 = \sim 12$ ). Non è ovviamente possibile sapere in anticipo quali dei 100 bambini manifesteranno questi effetti collaterali. L'NNH è un indicatore terapeutico dei trattamenti: maggiore è il suo valore, più sicuro è quel trattamento e viceversa. L'NNH può essere rappresentato anche graficamente, come viene proposto nel sito internet <http://www.nntonline.net/>.

## **Odds**

🚩 Odds è un termine intraducibile in italiano, mutuato dal mondo delle scommesse. Esprime il rapporto tra la probabilità che un evento si verifichi rispetto alla probabilità che esso non si verifichi (variabili binarie o dicotomiche, come per esempio maschi/ femmine, esposti/non esposti, malati/non malati, positivi/negativi a un test, presenza/ assenza di una complicanza). Se è superiore a 1 è più probabile che avvenga il successo (rispetto al non successo); se è inferiore a 1, viceversa, è più probabile che avvenga l'insuccesso; se è uguale a 1, i 2 eventi hanno la stessa probabilità. In caso di eventi rari l'odds ha un valore molto vicino a quello della corrispondente probabilità.

## Odds ratio (OR)

L'odds ratio è dato dal rapporto tra 2 odds. Se il valore dell'odds ratio è superiore a 1 significa che la probabilità che si verifichi l'evento considerato (per esempio una malattia) in un gruppo (per esempio tra gli esposti) è superiore rispetto a quella di un altro gruppo (per esempio tra i non esposti). Significato opposto ha un valore inferiore a 1, mentre se il valore è pari a 1 significa che non vi è differenza tra i gruppi. In caso di eventi rari l'odds ratio ha un valore molto vicino a quello del rischio relativo. L'odds ratio è una misura particolarmente utile negli studi caso-controllo come stima del rischio relativo, che in questo tipo di studi non può essere misurato direttamente. Nella tabella sottostante è riportato un esempio di come si valuterebbe l'effetto del trattamento con betabloccanti sulla comparsa di eventi cardiovascolari gravi in soggetti ipertesi, rispetto al trattamento con placebo.

Gruppo di trattamento	Eventi cardiovascolari		
	Sì (malato)	No (non malato)	totale
Betabloccanti (caso)	5	350	355
Placebo (controllo)	9	340	349
Totale	14	690	704

Per i pazienti trattati con betabloccanti l'odds di avere un evento è pari a  $5/350$ , quindi  $0,0143$ ; per quelli trattati con placebo l'odds è  $9/340=0,0265$ . L'odds ratio di avere un evento cardiovascolare nei pazienti trattati con betabloccanti rispetto a quelli trattati con placebo sarà dato dal rapporto  $0,0143/0,0265$ , ossia  $0,54$ . Questo significa che i pazienti trattati con betabloccanti hanno un odds di  $0,54$  volte inferiore rispetto ai trattati con placebo di avere un evento cardiovascolare grave (in questo caso, trattandosi di eventi rari, gli odds sono molto vicini alle probabilità, pari rispettivamente a  $0,0141$  e  $0,0258$ , e l'odds ratio molto vicino al rischio relativo, che in questo caso è  $0,546$ ).

## Placebo

Un placebo è un trattamento biologicamente inerte che viene somministrato ai pazienti di un gruppo di controllo in uno studio sperimentale, allo scopo di eliminare la possibilità che i soggetti che non ricevono un trattamento attivo siano influenzati dalle proprie aspettative e dunque di stimare l'effetto del principio attivo al netto degli effetti di natura psicologica legati all'atto terapeutico. Tale trattamento è indistinguibile (per colore, odore, sapore eccetera) dal trattamento di cui si sta indagando l'efficacia.

## Possibili disegni di studio

Dal punto di vista del disegno, **uno studio può essere:**

- **descrittivo**, se si limita a descrivere la frequenza di malattie, fattori di rischio, pratiche assistenziali o altro;
- **analitico**, se cerca di chiarire associazioni e nessi causali tra i diversi eventi;
- **esaustivo**, se prende in esame la totalità dei soggetti;
- **campionario**, se recluta un sottoinsieme dei soggetti stessi, ottenendo così delle stime in grado di conciliare le opposte esigenze della precisione e dell'economicità. Sono noti e usati molti disegni di campionamento, dal campione casuale semplice al campione stratificato, a più stadi eccetera;
- **primario**, che prevede la raccolta di informazioni originali;
- **secondario**, che fa cioè la revisione di uno o più studi primari.

Uno **studio primario** può essere:

- **OSSERVAZIONALE**, senza intervento esterno del ricercatore, che si limita a osservare gli eventi nei gruppi in studio;
- **SPERIMENTALE**, con intervento esterno da parte dello sperimentatore, che cerca di influenzare la frequenza degli eventi. Il classico studio sperimentale è lo studio randomizzato;
- **CONTROLLATO**, con almeno 2 gruppi di soggetti che subiscono interventi diversi;
- **NON CONTROLLATO**, quando manca un gruppo di controllo.

Uno **studio osservazionale** può essere:

- **trasversale**, se i fatti non si svolgono nel tempo, ma sono rilevati in un breve periodo che si assume istantaneo; c'è la possibilità di misurare la prevalenza ma non l'incidenza;
- **longitudinale**, se i fatti si svolgono nel tempo; c'è la possibilità di misurare l'incidenza o altri indici di frequenza nel tempo.

Uno **studio longitudinale** può essere:

- **di coorte**, o prospettico, in cui il reclutamento avviene sulla base dell'esposizione e si misura successivamente l'esito;
- **caso-controllo**, o retrospettivo, in cui il reclutamento avviene sulla base dell'esito e si misura a posteriori l'esposizione.

Uno **studio sperimentale** può essere:

- **randomizzato**, quando l'assegnazione ai trattamenti avviene per scelta casuale;

- quasi randomizzato, quando l'assegnazione ai trattamenti avviene con metodi o non completamente casuali o comunque facilmente decodificabili;
- non randomizzato, quando l'assegnazione ai trattamenti avviene secondo criteri non casuali.

Uno **studio randomizzato** può essere:

- parallelo, con almeno 2 gruppi di soggetti, ognuno dei quali riceve solitamente un solo trattamento;
- *crossover*, con un solo gruppo di soggetti che ricevono diversi trattamenti in momenti diversi;
- o avere altri disegni (fattoriale eccetera).

Uno **studio secondario** può essere una:

- revisione tradizionale;
- revisione sistematica, analisi della letteratura scientifica caratterizzata da una modalità esaustiva e riproducibile degli articoli rilevanti per rispondere a un preciso quesito clinico.

**Studio randomizzato:** Lo schema di base di uno studio randomizzato è caratterizzato dalle seguenti fasi:

Individuazione di un campione di soggetti da un gruppo più ampio, idealmente l'intera popolazione dei pazienti con la malattia in esame. I pazienti vengono poi selezionati secondo i criteri prefissati di inclusione ed esclusione dallo studio (eleggibilità).

I soggetti che hanno accettato di partecipare allo studio vengono divisi in 2 o più gruppi attraverso un procedimento casuale (randomizzazione). I diversi gruppi (bracci di trattamento) vengono quindi sottoposti ai rispettivi interventi. Nel caso di 2 gruppi, uno riceve il trattamento in esame (sperimentale) l'altro il trattamento di riferimento (il migliore disponibile, se già esistono altre terapie, altrimenti il placebo). Se necessario per preservare la validità dello studio né i clinici né i pazienti sono a conoscenza di quale gruppo sta prendendo il trattamento sperimentale e quale quello di controllo (cecità). I 2 gruppi di pazienti vengono poi sottoposti a interventi ulteriori, se necessario, e a controlli in modo identico.

Alla fine dello studio, così come programmato nel protocollo di studio iniziale, si analizzano i risultati. In assenza di errori sistematici, le differenze negli esiti in esame possono essere attribuite al tipo di intervento in sperimentazione.

**Revisione sistematica:** Una revisione sistematica esamina e sintetizza i dati provenienti da più studi primari prodotti su un certo quesito clinico. La ricerca delle fonti e l'analisi dei risultati sono appunto sistematici, cioè caratterizzati da un metodo esplicito e riproducibile. Se si decide che è appropriato e fattibile sulla base delle caratteristiche degli studi, si può procedere alla combinazione quantitativa dei dati secondo un processo chiamato metanalisi. La stima quantitativa combinata dei dati dei singoli studi di pazienti provenienti dai diversi studi ha il vantaggio di fornire una stima dell'effetto medio dell'intervento e di dare una idea precisa della forza dell'associazione. Una volta chiarito il quesito cui dare una risposta, il metodo di conduzione di una revisione sistematica prevede:

elaborazione di criteri di eleggibilità degli studi da includere nella revisione;

loro ricerca esaustiva nella letteratura;

valutazione qualitativa di ognuno di essi;

metanalisi, caratterizzata da:

- combinazione dei risultati in un statistica sintetica (determinazione della dimensione dell'effetto);
- valutazione dell'omogeneità della dimensione dell'effetto tra gli studi selezionati;
- preparazione di un elaborato finale di commento che, oltre a riportare i risultati, descriva anche i principali punti di forza e di debolezza.

Ogni metanalisi tratta i singoli lavori come uno studio tratta i singoli pazienti. Perciò misura l'effetto globale come una media ponderata degli effetti rilevati da ognuno dei lavori inclusi nella revisione, con coefficienti di ponderazione variabili secondo il metodo usato. Vi sono 2 diversi modelli di analisi: il modello a effetti fissi (*fixed effects*) e il modello a effetti casuali (*random effects*).

## **Potenza statistica**

➤ Si riferisce alla probabilità di mostrare come statisticamente significativa ➤ una differenza di effetto di un trattamento o di un'esposizione, quando questa realmente esiste. La potenza statistica dipende dal numero di eventi che si verificano in uno studio e dal numero di soggetti inclusi nel campione. E' il complemento a 1 della probabilità di errore di secondo tipo (potenza = 1-beta); quanto più uno studio è dotato di potenza, tanto meno probabile è il rischio di commettere un errore di secondo tipo, giudicando non significativa una differenza invece esistente.

## **Prevalenza**

➤ La prevalenza è la proporzione di persone facenti parte di una popolazione definita (per esempio una nazione) che in un momento determinato (per esempio oggi) risultano affette da una certa condizione (una malattia, per

esempio diabete; un parametro alterato, per esempio ipercolesterolemia). Il tasso di prevalenza, solitamente istantaneo, può anche essere esteso a periodi temporali diversi: giorni, settimane, mesi, anni. Per "prevalenza nella vita" si intende la proporzione di persone che, nell'intero corso della loro vita, presentano in un qualche momento il disturbo, il fenomeno o il valore preso in esame. In questo caso sarebbe più appropriato il termine incidenza o rischio. La prevalenza è una misura di frequenza adatta per lo studio di condizioni tendenzialmente croniche o protratte, mentre è meno adatta per lo studio di condizioni acute.

## Prova dell'ipotesi

Il percorso logico e metodologico per la verifica di un'ipotesi in ambito medico è riassumibile in alcuni punti. Si stabilisce innanzitutto un'ipotesi di partenza, la cosiddetta ipotesi nulla, che è sempre la più conservativa; per esempio: "il farmaco studiato ha un effetto medio pari a quello del farmaco di controllo". L'esperimento deve tentare di confutare questa ipotesi, dimostrando al contrario un'ipotesi alternativa per la quale "il farmaco studiato ha un effetto medio superiore al farmaco di controllo". Con la somministrazione dell'intervento studiato, se l'ipotesi di partenza è vera, non si dovrebbero osservare differenze tra i 2 gruppi: stessa proporzione di guariti, di migliorati o di peggiorati, stesse alterazioni dei parametri misurati, o comunque di qualsiasi effetto valutato. In realtà i valori misurati (in questo esempio 2 medie) quasi mai coincideranno tra i 2 gruppi. E' infatti poco probabile la perfetta uguaglianza di 2 valori che, anche solo per puro effetto del caso, tenderanno invece a essere leggermente diversi tra loro. Sarebbe perciò un errore, detto di primo tipo, rifiutare l'ipotesi nulla quando essa è vera; in altre parole attribuire al trattamento un effetto inesistente, frutto solo del caso. Come riconoscere allora gli effetti del caso da quelli veri, dovuti cioè a una reale efficacia dell'intervento analizzato? Il buon senso suggerisce che piccole differenze saranno più probabilmente frutto del caso, mentre grandi differenze dovranno insospettirci. Grazie a tecniche statistiche, è possibile associare a ogni differenza osservata una probabilità che tale differenza si verifichi. Si stabilisce convenzionalmente una soglia di probabilità, solitamente il 5 o l'1%, al di sotto della quale si prenderà la decisione di rifiutare l'ipotesi nulla. Questa probabilità è detta alfa, o livello di significatività, e rappresenta dunque il rischio, che si accetta a priori di correre, di cadere nell'errore di primo tipo, di assumere cioè come frutto di un trattamento una differenza che in realtà è frutto solo del caso. Il complemento a 1 di alfa, 1-alfa, si chiama "livello di protezione dall'errore di primo tipo". Per esempio per  $\alpha = 0,05$ , 1-alfa è  $= 0,95$  (95%). Se il ricercatore decidesse, per proteggersi dall'errore di primo tipo, di limitare al massimo il livello di alfa, correrebbe il rischio di commettere l'errore opposto: non rifiutare l'ipotesi nulla quando essa è falsa o, in parole più semplici, non vedere una differenza reale tra i 2 farmaci. Anche in questo caso si verificherebbe un errore, detto di secondo tipo. La probabilità di commettere questo errore è detta beta. Il suo complemento a 1, 1-beta, esprime la potenza del test, e in effetti misura il potere che un test ha di



svelare un effetto esistente. Per esempio per  $\beta = 0,2$ ,  $1-\beta$  è  $=0,8$  (80%).

Ricapitolando, esistono 2 tipi di errore, di segno opposto.

**L'errore di primo tipo** consiste nel rifiutare l'ipotesi nulla quando essa è vera (conseguenza: si attribuisce a un trattamento una differenza osservata che invece si è prodotta per caso).

**L'errore di secondo tipo** consiste nel non rifiutare l'ipotesi nulla quando essa è falsa (conseguenza: si attribuisce al caso una differenza osservata, che invece è dovuta al trattamento). Solitamente negli studi di buona qualità i ricercatori decidono il livello di alfa e di beta, e di conseguenza predeterminano la numerosità di soggetti da reclutare in ogni gruppo. Dopo l'esecuzione dell'esperimento, si misura sui dati ottenuti la probabilità, detta P, di ottenere i dati osservati, sotto ipotesi nulla: se il valore di P è inferiore al livello prefissato di alfa (per esempio  $P < 0,05$ ), si rifiuta l'ipotesi (attribuendo perciò le differenze osservate al trattamento), affermando che l'effetto osservato è statisticamente significativo, altrimenti non la si rifiuta, attribuendo allora le differenze osservate al caso. In entrambe le condizioni si è consapevoli della probabilità di errore della decisione, avendola fissata a priori.

Un altro modo per stabilire la significatività dell'effetto di un trattamento è **l'intervallo di confidenza**, che con una certa probabilità (pari al livello di confidenza, per esempio 95% o 99%) permette di essere nel giusto affermando che il vero valore del parametro studiato è compreso all'interno dell'intervallo stesso. I limiti di confidenza quantificano l'imprecisione o l'incertezza della stima di un certo parametro (per esempio aumento del rischio, rischio relativo, numero di casi da trattare) ricavata da un campione estratto dalla popolazione di cui si desidera conoscere il parametro. L'ampiezza dell'intervallo di confidenza dipende:

dalla numerosità del campione studiato; varia infatti proporzionalmente al variare della numerosità del campione; l'intervallo di confidenza dunque è legato alla potenza dello studio: sarà tanto più stretto, e di conseguenza la precisione della stima tanto maggiore, quanto più lo studio è potente;

dalla variabilità che il fenomeno studiato ha all'interno della popolazione; dal livello di confidenza prescelto: limiti di confidenza al 95%, a parità delle altre condizioni, sono più stretti rispetto ai limiti al 99%.

## Randomizzazione

➤ Assegnazione casuale dei soggetti inclusi in uno studio a uno dei gruppi di trattamento (disegno parallelo) o a una certa successione di trattamenti (disegno *crossover* ➤). La randomizzazione ha lo scopo di rendere tendenzialmente simili i gruppi per le loro caratteristiche, soprattutto per i fattori prognostici e di risposta alla terapia in esame, sia quelli noti e misurabili sia quelli ignoti. La randomizzazione permette l'applicabilità dei modelli probabilistici, sui quali si regge dal punto di vista metodologico l'intero studio.

In questo modo le differenze eventualmente osservate tra i gruppi di pazienti possono essere attribuite al trattamento.

## **Randomizzazione a cluster**

➤ E' una modalità di assegnazione a un trattamento con metodo casuale che invece di randomizzare singoli soggetti randomizza gruppi di pazienti.

## **Rapporto di verosimiglianza**

➤ E' il rapporto tra la frequenza con cui un certo valore di un test diagnostico si presenta tra i portatori di una certa malattia e la frequenza con cui lo stesso valore si presenta nei non affetti dalla malattia. Rappresenta un moltiplicatore che permette di passare dall'odd di malattia a priori all'odd di malattia a posteriori. Il rapporto di verosimiglianza del test positivo ha generalmente un valore superiore a 1, dunque incrementa la probabilità di malattia. Il rapporto di verosimiglianza del test negativo ha generalmente un valore inferiore a 1, dunque riduce la probabilità di malattia.

## **Rapporto di verosimiglianza negativo**

➤ E' il rapporto tra la frequenza con cui un individuo con la condizione/malattia indagata presenta un risultato negativo del test diagnostico e la frequenza con cui un individuo senza la condizione/malattia indagata presenta un risultato negativo dello stesso test. Il rapporto di verosimiglianza del test negativo ha generalmente un valore inferiore a 1, dunque riduce la probabilità di malattia.

## **Rapporto di verosimiglianza positivo**

➤ E' il rapporto tra la frequenza con cui un individuo con la condizione/malattia indagata presenta un risultato positivo del test diagnostico e la frequenza con cui un individuo senza la condizione/malattia indagata presenta un risultato positivo dello stesso test. Il rapporto di verosimiglianza del test positivo ha generalmente un valore superiore a 1, dunque incrementa la probabilità di malattia.

## **Revisione sistematica**

➤ Le revisioni sistematiche sono sintesi sistematiche delle conoscenze, basate sulla valutazione critica comparata della letteratura disponibile su un determinato argomento. Il termine sistematico si riferisce al fatto che la revisione deve essere pianificata come una vera e propria ricerca scientifica, attraverso la predisposizione di un protocollo che espliciti obiettivi, modalità di ricerca, reperimento, valutazione critica e sintesi di tutti gli studi eleggibili per la revisione. Nell'ambito di una revisione sistematica, se sussistono le condizioni di similarità fra tipi di pazienti e di trattamenti indagati, può essere realizzata una metanalisi ➤. Per evitare possibili distorsioni indotte da *bias* di

pubblicazione<sup>7</sup>, una revisione sistematica di buona qualità deve prevedere una ricerca esaustiva degli studi, anche se parzialmente pubblicati (per esempio *abstract* presentati a convegni) o non pubblicati (per esempio tesi di laurea, rapporti di ricerca eccetera).

## Riduzione del rischio assoluto (ARR)

<sup>7</sup>La riduzione del rischio assoluto (o riduzione assoluta del rischio) rappresenta la differenza nel rischio assoluto<sup>7</sup> fra gruppo di controllo<sup>7</sup> e gruppo sperimentale. Questa misura è impiegata quando il rischio nel gruppo sperimentale è inferiore al rischio nel gruppo di controllo e viene calcolata sottraendo il rischio assoluto<sup>7</sup> nel gruppo sperimentale dal rischio assoluto nel gruppo di controllo. Si tratta di un parametro che non dà informazioni sull'aumento proporzionale fra i 2 gruppi; per quest'ultima informazione bisogna calcolare la riduzione del rischio relativo<sup>7</sup>. Un valore superiore a 0 indica effettivamente un beneficio, mentre un valore negativo ha il significato di un rischio; un valore pari a 0 indica l'indifferenza tra i 2 trattamenti (esposizioni). Nella tabella sottostante è riportato un esempio di come si valuterebbe l'effetto del trattamento con betabloccanti sugli eventi cardiovascolari gravi in pazienti ipertesi, rispetto al trattamento con placebo<sup>7</sup>.

Gruppo di trattamento	Eventi cardiovascolari		
	sì	no	totale
Betabloccanti	5	350	355
Placebo	9	340	349
Totale	14	690	704

Il rischio assoluto di evento cardiovascolare grave nei trattati con placebo è pari a 2,58% (9/349), il rischio assoluto nei trattati con betabloccanti è dell'1,41% (5/355). La riduzione del rischio assoluto si calcola come  $2,58\% - 1,41\% = 1,17\%$ . L'inverso della riduzione del rischio assoluto è il numero di casi da trattare (NNT, in questo caso 85).

## Riduzione del rischio relativo (RRR)

<sup>7</sup>La riduzione del rischio relativo è la riduzione proporzionale del rischio tra i soggetti che fanno parte del gruppo sperimentale rispetto a quello presente tra i soggetti del gruppo di controllo<sup>7</sup>. La riduzione del rischio assoluto<sup>7</sup> (differenza di rischi tra i 2 gruppi) viene rapportata al rischio del gruppo di controllo ed espressa in termini percentuali. Un valore superiore a 0 indica una riduzione di rischio, mentre un valore negativo ha il significato di un incremento di rischio; un valore pari a 0 indica l'indifferenza fra i 2 trattamenti (esposizioni). Nella tabella sottostante è riportato un esempio di come si valuterebbe l'effetto del

trattamento con betabloccanti sugli eventi cardiovascolari gravi in pazienti ipertesi, rispetto al trattamento con placebo<sup>7</sup>.

Gruppo di trattamento	Eventi cardiovascolari		
	sì	no	totale
Betabloccanti	5	350	355
Placebo	9	340	349
Totale	14	690	704

La riduzione del rischio relativo nei pazienti trattati con betabloccanti si calcola come  $(R_c - R_t) / R_c$ , dove  $R_c$  è il rischio assoluto di evento cardiovascolare grave nei trattati con placebo (rischio nei controlli, in questo caso:  $9/349$ , 2,58%) e  $R_t$  il rischio assoluto di evento cardiovascolare grave nei trattati con betabloccanti (rischio nei trattati, in questo caso:  $5/355$ , 1,41%). La riduzione del rischio relativo è quindi  $(2,58\% - 1,41\%) / (2,58\%) = 45,4\%$ . Si può dunque affermare che l'assunzione della terapia riduce il rischio di evento cardiovascolare grave del 45% circa.

### Rischio assoluto (AR)

<sup>7</sup>Il rischio assoluto è dato dal numero di nuovi casi di malattia in un certo periodo di tempo, rapportato al numero di soggetti a rischio all'inizio del periodo. Indica la probabilità che un individuo presenti l'evento studiato nel corso di un periodo di tempo specificato. Nel caso di 2 gruppi, va calcolato distintamente nel gruppo dei trattati e nel gruppo dei non trattati. Il suo valore può variare da 0 a 1 (o in termini percentuali da 0% a 100%).

### Rischio relativo (RR)

<sup>7</sup>Il rischio relativo, in inglese *risk ratio*, è il rapporto tra il rischio assoluto<sup>7</sup> di un evento<sup>7</sup> (solitamente sfavorevole) in 2 gruppi (trattati e non trattati, esposti e non esposti). Il rischio relativo si valuta negli studi di coorte<sup>7</sup> o negli studi controllati e randomizzati<sup>7</sup>, ed esprime la forza dell'associazione fra il trattamento (o l'esposizione) e la malattia. Un valore superiore a 1 indica un aumento di rischio legato al trattamento o all'esposizione, mentre un valore inferiore a 1 ha il significato di un effetto benefico o protettivo; un valore pari a 1 indica l'indifferenza tra i 2 trattamenti (esposizioni). Se si volesse valutare l'effetto del trattamento con betabloccanti sugli eventi cardiovascolari gravi in pazienti ipertesi, rispetto al trattamento con placebo<sup>7</sup>, sulla base dei dati riportati nella tabella sottostante il rischio relativo si calcola come:  $1,4\% / 2,6\% = 0,54$ .

Gruppo di trattamento	Eventi cardiovascolari		
	sì	no	totale
Betabloccanti	1,4% (5x100/355)	350	355
Placebo	2,6% (9x100/349)	340	349

## Sensibilità

La sensibilità di un test diagnostico indica la frequenza con cui quel test dà un risultato positivo nei soggetti con la malattia indagata. Un test molto sensibile risulterà positivo in un numero molto alto di casi nei quali la malattia è effettivamente presente (bassa proporzione di falsa negatività); un test poco sensibile risulterà positivo in un numero molto basso di casi nei quali la malattia è effettivamente presente (alta proporzione di falsa negatività).

## Specificità

La specificità di un test diagnostico indica la frequenza con cui quel test dà un risultato negativo nei soggetti senza la malattia indagata. Un test molto specifico risulterà negativo in un numero molto alto di casi nei quali la malattia è effettivamente assente (bassa proporzione di falsa positività); un test poco specifico risulterà positivo in un numero molto basso di casi nei quali la malattia è effettivamente assente (alta proporzione di falsa positività).

## Statisticamente significativo

L'espressione "statisticamente significativo" (o semplicemente "significativo" nel testo di CE) viene usata per indicare una bassa probabilità che le differenze osservate nei campioni studiati siano erroneamente attribuite al caso (errore di primo tipo). Convenzionalmente si fa riferimento, come valore soglia, al livello di significatività del 5% (alfa <0,05): ciò significa che la probabilità calcolata a posteriori ( $P$ ) che il risultato osservato si verifichi per caso, se fosse vera l'ipotesi di partenza di non differenza tra i 2 interventi/trattamenti (ipotesi nulla), è inferiore al 5%. Il livello di significatività scelto come soglia da chi conduce uno studio può essere anche inferiore o superiore e dipenderà dal grado di certezza con cui si vuole escludere che l'eventuale differenza trovata sia attribuibile al caso.

## Studio caso-controllo

E' un tipo di disegno di studio che confronta un gruppo di persone che hanno manifestato un evento/malattia (casi) con un gruppo di persone che non hanno manifestato lo stesso evento/malattia (controlli) rispetto alla presenza/assenza di esposizione a fattori di rischio sospettati di poter determinare quella malattia. In questo tipo di disegno, necessariamente retrospettivo, i 2 gruppi devono venire assemblati in modo da essere simili per la maggior parte delle

variabili rilevanti (sesso, età, livello socio-economico eccetera) tranne che per la presenza della malattia. I vantaggi degli studi caso-controllo sono la relativa rapidità di esecuzione e il costo non elevato. Inoltre, questo tipo di disegno permette di studiare fattori di rischio che, per motivi etici o anche solo organizzativi, non sia possibile valutare sperimentalmente all'interno di uno studio prospettico. Tra gli svantaggi, la difficoltà nella scelta di controlli adeguati, il *bias* legato al ricordo di eventi passati.

### **Studio con disegno crossover**

Studio sperimentale nel quale i soggetti arruolati sono sottoposti a entrambi i trattamenti in successione (per esempio prima il farmaco A, poi il farmaco B). In tal modo ogni soggetto fa da controllo di se stesso. I pazienti possono essere randomizzati a diverse sequenze di trattamento (prima A poi B oppure prima B poi A); sono possibili anche disegni più complessi (più di 2 farmaci, molteplicità di periodi e sequenze). Il momento in cui i partecipanti del gruppo A passano ad assumere il trattamento del gruppo B e viceversa è detto *crossover*, cioè incrocio. Il disegno *crossover* è possibile solo per alcune condizioni, come le malattie croniche non guaribili; tra un periodo terapeutico e l'altro è necessaria la sospensione del farmaco per evitare interferenze fra i diversi trattamenti (*washout*).

### **Studio controllato e randomizzato**

Lo studio controllato e randomizzato (studio randomizzato nel testo di CE) è uno studio prospettico nel quale i partecipanti sono reclutati e assegnati in modo casuale (randomizzato) a 2 o più gruppi. Nel caso più comune dello studio parallelo con 2 gruppi (o bracci) uno, sperimentale, riceve l'intervento di cui si vuole valutare l'efficacia l'altro, il gruppo di controllo, riceve il trattamento di controllo (trattamento standard, nessun trattamento o un placebo). È il disegno di studio che riesce meglio a eliminare il *bias* di selezione e a garantire la confrontabilità dei gruppi. Per motivi etici uno studio randomizzato può essere effettuato soltanto per trattamenti di cui sia incerta l'efficacia (principio di incertezza). Nello studio randomizzato acquistano particolare rilievo: la tecnica utilizzata per la randomizzazione (dalla semplice estrazione da un'urna, all'uso di tabelle di numeri casuali o di programmi computerizzati per la generazione di tali numeri); la tecnica utilizzata per mantenere nascosta la sequenza di randomizzazione (*allocation concealment*); la cecità nei vari livelli; l'analisi dei dati secondo il principio dell'intenzione al trattamento.

### **Studio di coorte**

È un disegno di studio non sperimentale realizzato su uno o più gruppi di soggetti (coorti). All'interno di una coorte i soggetti hanno in comune, al momento dell'arruolamento, una o più caratteristiche. Lo studio di coorte può essere sia retrospettivo (su dati storici) sia prospettico. Nel caso di assenza di un gruppo di controllo, nell'unica coorte considerata si valuta nel tempo

l'incidenza degli eventi di interesse (esito). Lo studio di coorte controllato è realizzato mettendo a confronto almeno 2 coorti di soggetti, definiti dalla presenza o assenza (o differente intensità) di un'esposizione. Le coorti vengono seguite nel tempo per valutare se e in che misura esse differiscano in termini di frequenza di sviluppo di una malattia/condizione (esito), in funzione della presenza/assenza dell'esposizione. Le misure dell'effetto sono solitamente i rischi o incidenze cumulative, oppure i tassi di incidenza e i relativi rapporti tra rischi e tra incidenze (rischi relativi). Questo disegno può essere poco vantaggioso per malattie rare o con lunga fase di latenza, richiedendo in tal caso grandi quantità di risorse.

### **Studio di coorte prospettico**

Studio di coorte che segue i partecipanti nel tempo; il reclutamento è nel presente, il follow up nel futuro. E' più affidabile dello studio di coorte retrospettivo (su dati storici), che invece ricostruisce, attraverso documenti appositi, la storia dei soggetti (reclutamento e follow up nel passato).

### **Studio in aperto**

Studio sperimentale nel quale sia i partecipanti sia i somministratori degli interventi sono a conoscenza del tipo di trattamento che viene utilizzato in ogni soggetto. Ciò può condizionare l'interpretazione dei risultati. In questo genere di studi si consiglia che i valutatori degli esiti siano in condizione di cecità.

### **Studio in cieco**

Studio sperimentale nel quale ogni singolo partecipante non ha la possibilità di capire quale tipo di trattamento stia seguendo (studio in singolo cieco). Se anche i somministratori degli interventi sono all'oscuro del tipo di trattamento del singolo paziente, si parla di studio in doppio cieco. Uno studio, infine, è detto in triplo cieco se anche chi analizza i dati raccolti e gli esiti dello studio non è a conoscenza del trattamento del singolo partecipante.

### **Studio sperimentale**

E' un tipo di studio in cui lo sperimentatore assegna in modo attivo i partecipanti all'esposizione (o all'eliminazione) di un fattore di rischio, o a un trattamento di cui si voglia valutare l'efficacia o gli effetti collaterali, confrontandoli poi con un gruppo di controllo costituito da soggetti non esposti a quello stesso fattore. Per motivi etici o di fattibilità gli studi epidemiologici sono generalmente di tipo non sperimentale, cioè basati su esposizioni ai fattori di rischio già esistenti.

### **Studio trasversale**

E' un disegno di studio, detto anche di prevalenza, in cui una popolazione definita viene esaminata in un preciso istante (o comunque in un tempo

sufficientemente breve da poter essere considerato quasi istantaneo) al fine di misurare la presenza di una malattia, dell'esposizione a un particolare fattore eziologico o di qualsiasi altra condizione. Lo studio trasversale può essere considerato una fotografia istantanea del gruppo di persone esaminate e non richiede che i soggetti vengano seguiti nel tempo (né retrospettivamente né prospettivamente). Le misure utilizzate sono pertanto la prevalenza o i rapporti tra prevalenze nel caso di studi controllati. Il maggiore vantaggio dello studio trasversale è di essere generalmente rapido ed economico. Data l'elevata probabilità di *bias* e l'impossibilità di valutare correttamente i nessi causali tra esposizioni ed esiti, questo tipo di studio viene spesso utilizzato in prima battuta come generatore di ipotesi per studi dal disegno più robusto.

### **Tasso di morbilità**

Il termine tasso equivale al termine inglese *rate*. In una determinata popolazione di riferimento si esprime come rapporto tra il numero di casi di una certa malattia e il numero di soggetti esposti in un periodo di tempo definito, tenendo conto del tempo di esposizione di ogni singolo soggetto e dunque del tempo totale di esposizione (tempo per persone). Negli studi su grandi popolazioni viene assunto come denominatore la popolazione media nel periodo.

### **Tasso di mortalità**

Il termine tasso equivale al termine inglese *rate*. Un tasso grezzo di mortalità è il numero di decessi che si verifica in un periodo di tempo definito, rapportato a una popolazione di dimensione nota (media nel periodo considerato). In genere si esprime come numero di decessi per 1.000 o 100.000 soggetti della popolazione studiata.

### **Validità**

Questo termine può essere usato con riferimento a uno strumento di misurazione, nel qual caso indica il grado di affidabilità della misura, o a uno studio: in questo caso si deve distinguere tra validità interna e validità esterna (o generalizzabilità). La validità interna di uno studio si riferisce al rispetto delle regole metodologiche che hanno come fine la riduzione al minimo di errori sistematici nel disegno, nella conduzione e nell'analisi dello studio stesso; essa fa pertanto riferimento al rigore metodologico complessivo dello studio. La validità esterna si riferisce invece alla trasferibilità, generalizzabilità e applicabilità dei risultati dello studio a popolazioni e soggetti di interesse nella pratica quotidiana, ed è in larga misura dipendente dalle caratteristiche del campione studiato. Quanto maggiore è la somiglianza con i pazienti cui vanno applicati i risultati dello studio, tanto maggiore è la validità esterna dello studio. Oggi si tende a conciliare queste 2 dimensioni della validità, realizzando studi multicentrici di grandi dimensioni.



## **Valore di P**

Il valore di P esprime la probabilità di ottenere i risultati osservati, o più estremi ancora, nel caso sia vera l'ipotesi nulla: se il valore di P è inferiore al livello prefissato di alfa (per esempio  $P < 0,05$ , 1 su 20) si rifiuta l'ipotesi (attribuendo le differenze osservate al trattamento), altrimenti non la si rifiuta (attribuendo le differenze osservate al caso). Il valore di P, esito dei test statistici (per esempio il test t di Student, il test zeta e il test Chi quadro) misura la probabilità di commettere un errore (di primo tipo) nel considerare l'effetto osservato come attribuibile al trattamento e non al caso. Se questa probabilità è inferiore a 1 su 20 ( $P < 0,05$ ) allora il risultato è, convenzionalmente, considerato come statisticamente significativo.

## **Valore predittivo negativo**

Indica la probabilità che, dato un soggetto con un test diagnostico dall'esito negativo, risulti corretta l'attribuzione di quel soggetto alla categoria dei non affetti da una certa malattia. Da non confondere con la specificità.

## **Valore predittivo positivo**

Indica la probabilità che, dato un soggetto con un test diagnostico dall'esito positivo, risulti corretta l'attribuzione di quel soggetto alla categoria degli affetti da una certa malattia. Da non confondere con la sensibilità.

## **Vero negativo**

Un vero negativo è un risultato corretto di un test diagnostico, tale per cui un soggetto non affetto dalla condizione/ malattia indagata ha un risultato negativo del test diagnostico effettuato per rilevarne la presenza.

## **Vero positivo**

Un vero positivo è un risultato corretto di un test diagnostico, tale per cui un soggetto affetto dalla condizione/malattia indagata ha un risultato positivo del test diagnostico effettuato per rilevarne la presenza.