

Clementino Stefanetti

c.stefanetti@tin.it



**Statistica
per il MMG**

INDICE

Cap 1 IL CAMPIONAMENTO	4	Cap 3 GLI INDICI DI TENDENZA CENTRALE	52
Studi osservazionali		Media aritmetica semplice	
Trasversali		Media aritmetica ponderata	
Caso-controllo		Media geometrica	
Longitudinali (coorti)		Media armonica	
Studi sperimentali		Moda	
		Mediana	
Il BIAS		Cap 4 GLI INDICI DI DISPERSIONE	58
di selezione			
di informazione		Varianza	
di confondimento		Deviazione standard	
		Errore standard	
Il metodo random		Cap 5 L'INTERVALLO DI CONFIDENZA	62
Randomizzazione semplice			
Randomizzazione ristretta		Significato dell'intervallo di confidenza	
Blocchi permutati		Uso dell'intervallo di confidenza	
Stratificazione		Cap 6 SIGNIFICATO DI SIGNIFICATIVO E POTENZA DEL TEST	68
Minimizzazione			
Disegni con consenso		I significati di Significativo	
Randomizzazione telefonica		Errori comuni dell'applicazione del test T	
Studi in doppio cieco		Potenza del test: falsi positivi e falsi negativi	
Metanalisi		La dimensione dell'errore	
		La differenza che si vuole rilevare	
Cap 2 STATISTICA EPIDEMOLOGICA	40	La numerosità campionaria	
		Formule per il calcolo della Potenza	
Prevalenza e incidenza		Cap 7 LA RICERCA CON MEDLINE IN MEDICINA	76
Tassi e quozienti			
Tasso di prevalenza			
Tasso di incidenza			
Tasso di morbilità			
Tasso di letalità			
Tasso di mortalità			
Numeri indice			
NNT e NNH			
RCT e metanalisi			
Diagrammi di odds ratio			
Weighted means difference (WMD)			
Test di screening			
Spectrum BIAS			

Introduzione

È molto probabile che alcuni colleghi, leggendo il titolo di questo lavoro, aggrottino la fronte e passino oltre, nel timore di essere coinvolti in qualche cosa di noioso o quanto meno di astruso. Quando un medico, che vive quotidianamente la sua professione in un ambulatorio frequentato da ammalati, si trova di fronte improvvisamente a termini quali “Deviazione Standard” o “Odds Ratio” la prima reazione che ha è di istintiva repulsione, come sempre accade quando affrontiamo argomenti molto distanti dal nostro modo di vivere e di ragionare.

È comprensibile, ma è profondamente sbagliato.

La medicina di Ippocrate e di Galeno ha subito nel corso dei secoli un processo di “raffinatura” che è passato anche attraverso le forche caudine della matematica e della statistica, ed alle soglie del terzo millennio non è più concepibile un medico di famiglia che non sia in grado quanto meno di orientarsi tra i concetti più elementari di queste scienze esatte.

La spinta a scrivere questo libretto è stata esattamente questa: una “sfida” alla conoscenza, un impegno a tramutare in esempi pratici e comprensibili quelli che di primo acchito possono apparire come aridi simbolismi, una rapida carrellata, vista con gli occhi di un medico e destinata ai medici, delle basi statistiche fondamentali che sono l'indispensabile chiave di lettura dei lavori scientifici.

Se si avrà la costanza e l'umiltà di sedersi nuovamente sui banchi scolastici, al termine della lettura avremo imparato assieme concetti quali NNT, Odds Ratios, Intervalli di Confidenza, Potenza o Deviazione Standard, che devono far parte della cultura medica con la stessa dignità della farmacologia o della clinica, permettendo al clinico non solamente la conoscenza ma anche la comprensione e la valutazione dei trials e delle meta-analisi che costituiscono il bagaglio quotidiano dell'aggiornamento.

Clementino Stefanetti MMG (Casorezzo - MI)

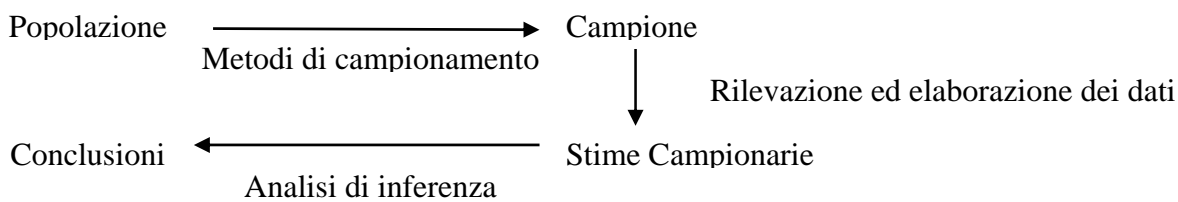
IL CAMPIONAMENTO

Il termine **popolazione** è l'insieme di tutti gli elementi (unità statistiche) con un carattere distintivo comune che è appunto oggetto di studio. Ad esempio, la popolazione dei soggetti affetti da ipertensione comprende tutte le persone che hanno l'ipertensione.

Poiché è impensabile che si possa condurre una ricerca biologica direttamente sull'intera popolazione si prenderà un **campione** rappresentativo della popolazione. Il campione non è altro quindi che un sottoinsieme della popolazione globale.

Inoltre un'unità statistica può far parte contemporaneamente di più popolazioni a seconda delle sue caratteristiche: un individuo può essere o no affetto da ipertensione, fumatore o meno, diabetico o non diabetico, grasso o magro, maschio o femmina, e via dicendo.

In generale, quindi, una popolazione statistica dipende dalla variabile oggetto di indagine e viene definita dal ricercatore che deve stabilire un criterio oggettivo per assegnare o no un elemento alla popolazione che interessa. Le ricerche vengono quindi programmate con lo scopo di arrivare a conclusioni generali, valide per tutte le unità statistiche.



Dal momento che il campione rappresenta l'intera popolazione, esso deve fornire garanzie circa l'affidabilità dei risultati che saranno generalizzati a tutta la popolazione.

In termini statistici **il campione deve essere rappresentativo**, e per verificarsi questa condizione la sua scelta deve essere affidata al caso. Un campione non casuale non può risultare rappresentativo. I concetti di casualità e rappresentatività del campione sono quindi legati tra loro, anche se non sono sinonimi: una scelta casuale sulla popolazione è una condizione necessaria per raggiungere la rappresentatività, ma non è una condizione sufficiente per ottenerla. Può capitare, ad esempio, che nel saggiare un farmaco contro l'ipertensione si debba escludere dal gruppo degli ipertesi le donne in gravidanza o i cardiopatici. Questi criteri di esclusione, ovviamente, limitano la rappresentatività della popolazione.

La scelta del campione e, come vedremo poi, la sua numerosità sono i due parametri più importanti per condurre un buon studio.

La differenza esistente tra le caratteristiche del campione e quelle della popolazione di riferimento viene definita **errore di campionamento**. Difficilmente quest'errore può essere misurato. Tale errore può derivare da:

- 1) un errore nella selezione dei soggetti
- 2) casualità

1) Una selezione è viziata quando coinvolge un sottoinsieme non rappresentativo della popolazione. Il vizio può però essere evitato o ridotto adottando determinate modalità di reclutamento. Ad esempio, se il campione è formato da individui sani, questi lo devono essere effettivamente.

Per verificare la reale appartenenza di un soggetto a un campione si usa un controllo **in doppio cieco**: due ricercatori classificano il campione ognuno per conto suo.

Le unità con riassegnazione difforme vengono scartate dal campione. Se il numero di attribuzioni non concordanti supera il 10% delle unità è opportuno effettuare un nuovo campionamento.

2) L'errore legato al caso non è invece eliminabile, ma esistono tecniche statistiche per definirne, in modo probabilistico, la dimensione.

STUDI SUL RISCHIO

Per studiare i fattori di rischio ci sono due metodi.

Il metodo osservazionale e il metodo sperimentale. Il primo è indiretto: lo sperimentatore si limita ad osservare gli eventi senza agire su di essi. Il metodo sperimentale è diretto: lo sperimentatore compie interventi diretti, manipolando le condizioni della ricerca al fine di verificare l'efficacia delle misure adottate.

Classificazione degli studi sul rischio

- Studi osservazionali: analitici (trasversali, caso-controllo, longitudinali)
- Studi sperimentali: clinici, preventivi

Studi osservazionali analitici

Utilizzano dati pervenuti attraverso studi già esistenti (schede di decesso o registri ospedalieri, statistiche correnti): questi dati vengono opportunamente considerati e confrontati con i risultati ottenuti da studi analoghi. Lo scopo è quello di esaminare i vari fattori di rischio in rapporto a una determinata malattia al fine di fornire delle ipotesi sulle sue possibili cause.

Gli studi analitici hanno come principale obiettivo la dimostrazione delle relazioni esistenti tra le diverse componenti che emergono dalla ricerca (frequenza dell'evento, fattori di rischio, fattori di "confondimento").

a) Studi trasversali

Gli studi trasversali sono importanti per formulare ipotesi eziologiche sul rapporto causale tra malattie e fattori di rischio, soprattutto nel caso di malattie croniche non rare.

Inizialmente occorre definire il problema, oggetto di studio, che predispone alla malattia e la malattia che si intende studiare. Poi occorre definire il campione rappresentativo della popolazione.

La selezione del campione è il problema più delicato.

Il tipo più semplice è il *campionamento casuale semplice*, nel quale ogni persona ha la stessa probabilità di essere selezionata. Il modo è quello di assegnare un numero ad ogni persona partendo da 1,2,3 ecc. I numeri vengono poi scelti in modo casuale usando una tavola dei numeri casuali fino a raggiungere il numero del campione desiderato.

Un altro modo è il *campionamento casuale stratificato* che si ottiene suddividendo la popolazione in sottogruppi in base a caratteristiche importanti come età, sesso, condizione socio-economica ed estraendo un campione casuale da ogni sottogruppo. Se la percentuale di soggetti che costituiscono i diversi strati è la stessa che si trova nella popolazione oggetto di studio (es. il gruppo di età compresa tra 50 e 70 rappresenta il 25% della popolazione e il 25% è presente nel campione per quella fascia di età), allora tutti gli strati saranno equamente rappresentati dal punto di vista numerico, nel campione selezionato. Questa proporzionalità è consigliabile in quanto semplifica l'elaborazione dei dati.

Il campionamento a grappolo o cluster richiede:

- la suddivisione della popolazione in sottogruppi, o cluster, non necessariamente omogenei come gli strati.
- L'estrazione di un campione casuale dei cluster.
- La selezione di tutte le persone o di un campione casuale di persone in ogni cluster.

Il *campionamento sistematico* consiste nel decidere all'inizio quale frazione della popolazione debba essere studiata, esempio la metà o un quinto, e che venga compilato un elenco della popolazione in base a un certo ordine come in una rubrica telefonica. Poi partendo dall'inizio della lista viene scelta una persona ogni due o ogni dieci, in base all'ampiezza dell'intervallo deciso precedentemente. Per decidere il tipo di intervallo, il ricercatore deve sapere che l'intervallo non corrisponda a una distribuzione ricorrente nella popolazione.

Consideriamo cosa accadrebbe se la popolazione fosse costituita da una serie di coppie sposate in cui il nome del marito compare sempre per primo. Selezionando una persona ogni quattro, il campione risulterebbe

costituito solo da uomini, nel caso in cui si iniziasse dal primo nominativo o dal terzo, e solo da donne se si partisse dal secondo o dal quarto.

Consideriamo anche i vizi di fondo che possono sfuggire ai meno esperti. Il campionamento di nuclei familiari rappresenta un buon esempio. Se l'intervistatore che arriva non trova nessuno in casa, dovrebbe ripassare piuttosto che provare a suonare alla casa accanto, perché vi possono essere delle differenze fra i nuclei familiari i cui componenti rimangono a casa durante il giorno e quelli in cui ciò non accade. Allo stesso modo un intervistatore non dovrebbe iniziare sempre dalla prima abitazione di un isolato, perché le persone che vivono in case d'angolo potrebbero avere caratteristiche diverse rispetto a quelle che vivono a metà isolato.

Una volta definito il campione si devono raccogliere le informazioni necessarie. La presenza di una malattia può essere determinata in molti modi. Per esempio, in una piccola città, tutti o quasi tutti i casi di una specifica malattia possono essere identificati contattando i medici del luogo e consultando i registri ospedalieri o addirittura visitando direttamente tutti i residenti.

La presenza di possibili fattori causali sui quali si indaga può essere determinata anche attraverso esami e misurazioni appropriate. Per esempio, per valutare il ruolo di fattori causali attivi per via respiratoria, il numero di sigarette fumate può essere determinato attraverso un'intervista, e i livelli di inquinamento atmosferico in luoghi di residenza o di lavoro possono essere misurati con l'ausilio di apparecchiature.

Si passa poi all'analisi dei dati. Il metodo più seguito per la valutazione dei dati in uno studio trasversale consiste nel suddividere la popolazione secondo i fattori presi in esame e nel confrontare la prevalenza della malattia in ogni sottogruppo. Se si desidera studiare l'associazione fra neoplasia polmonare e abitudine al fumo in un gruppo di uomini di età media, si può dividere il gruppo in classi appropriate di fumatori; per esempio: nessuna sigaretta, fino a dieci sigarette al giorno, da dieci a venti, più di venti sigarette al giorno, ecc. Il tasso di prevalenza di neoplasia polmonare viene quindi determinato per ogni sottogruppo, e i tassi vengono poi confrontati fra loro.

L'interpretazione degli studi trasversali dimostra la presenza, o l'assenza, di un rapporto fra la variabile (o le variabili) in esame e il verificarsi di una malattia. Il calcolo dei casi di malattia accertata, a differenza di quelli riguardanti lo sviluppo della malattia, come negli studi di coorte, impone molta cautela, poiché può fornire dati non rappresentativi di tutti i casi di quella malattia.

Consideriamo per esempio la cardiopatia coronarica. Una delle sue più importanti manifestazioni è la morte improvvisa. In uno studio di prevalenza, i casi di cardiopatia coronarica che svilupparono come prima manifestazione la morte improvvisa non vengono considerati, in quanto l'intervallo di tempo in cui la malattia può essere riconosciuta è estremamente breve, e sarebbe inoltre eccezionale che il decesso capitasse proprio nel momento in cui l'individuo viene visitato nell'ambito dello studio. Partendo da questo esempio estremo è facile comprendere come la durata molto breve di una malattia, sia per rapido decesso sia per pronta guarigione, riduca la probabilità di individuarla mediante uno studio di prevalenza eseguito solo una volta.

Come logica conseguenza, ne deriva che i casi di lunga durata sono sovrastimati in uno studio trasversale.

Quando in uno studio di prevalenza prendiamo in considerazione la durata di malattia di una persona, è importante fare una piccola digressione e sottolineare che esistono due proprietà fondamentali di una malattia che si riflettono sulla sua prevalenza. Il primo è quanti casi di una malattia si sviluppano nell'unità di tempo, o incidenza; l'altro è quanto a lungo la malattia è presente, cioè la sua durata. In condizioni ideali di stabilità, in cui l'incidenza e la durata di una malattia rimangano costanti in un periodo di tempo, il rapporto fra prevalenza, incidenza e durata può essere espresso, quando la prevalenza sia inferiore al 10%, mediante una semplice equazione matematica:

Prevalenza = Incidenza x durata media

Così, se una delle tre misure è sconosciuta, la si può calcolare conoscendo le altre due, posto che si realizzi una condizione ideale di stabilità, come abbiamo detto precedentemente.

I casi di prevalenza possono diventare non rappresentativi di tutti i casi se alcuni fra i malati abbandonano la comunità. Alcune delle persone malate possono essere ricoverate altrove o recarsi in città dotate di migliori centri di cura e quindi sfuggire alle procedure locali di controllo.

Quando si interpretano i risultati di uno studio trasversale, bisogna fare attenzione a non assegnare un'inesistente sequenza temporale a un'associazione fra una caratteristica o un altro fattore e la malattia.. Se si scopre, per esempio, che i malati di cancro sono più ansiosi o hanno problemi emotivi più frequentemente dei soggetti sani in una stessa popolazione, non si può sostenere che l'ansia preceda il cancro. Dopo tutto, i malati di cancro hanno molte buone ragioni per essere preoccupati o per presentare dei disturbi emotivi. Non vi sarà alcun dubbio, invece sulla sequenza temporale tra cancro e altre caratteristiche quali il colore degli occhi, il gruppo sanguigno o l'esposizione della madre a radiazioni.

b) Studi caso-controllo

Gli studi caso-controllo sono molto simili agli studi trasversali se i casi considerati sono quelli esistenti in un determinato momento; oppure sono affini a quelli di coorte se i casi sono quelli nuovi o "incidenti", accumulatisi in un intervallo di tempo. Tuttavia, poiché negli studi caso-controllo viene esaminato un minor numero di soggetti, più facilmente reperibili, essi sono utilizzati molto più spesso. Fra gli studi analitici, solitamente rappresentano il primo approccio per valutare l'associazione tra una particolare caratteristica individuale o un fattore ambientale e una malattia.

Come viene condotto uno studio caso-controllo. Identificazione e raccolta dei casi. Precisati gli obiettivi e i metodi della ricerca, il primo passo da compiere in uno studio caso-controllo è l'identificazione dei casi, cioè delle persone malate che si vogliono studiare.

Come si è già accennato a proposito degli studi trasversali, è importante stabilire criteri per la diagnosi e l'inclusione dei casi nello studio e descrivere questi criteri dettagliatamente quando si riportano i risultati. In genere è bene richiedere un accertamento scrupoloso e documentato della malattia, anche se ciò potrà comportare l'esclusione di alcuni casi e quindi una riduzione del campione. Pertanto, in uno studio sulla calcolosi renale, è buona norma includere tutti gli individui in cui la presenza di calcoli sia documentata radiologicamente o mediante rimozione chirurgica e non solo in base alla presenza di coliche renali. Prendendo in considerazione casi non sufficientemente documentati, il ricercatore corre il rischio di diluire il gruppo dei casi con individui non realmente affetti dalla malattia, riducendo così le possibilità di evidenziare eventuali differenze fra il gruppo dei casi e il gruppo dei controlli.

I casi possono essere identificati mediante una ricerca condotta su un'intera comunità, il che risulta molto più facile qualora sia disponibile un registro di quella data malattia. Spesso i casi vengono accertati in uno o più ospedali, cliniche, centri medici e, più recentemente, consultando le cartelle dei medici. La selezione del gruppo si limiterà ai casi diagnosticati per la prima volta in un intervallo di tempo definito. Per esempio, si può decidere di studiare tutti i casi ben documentati di calcolosi renale scoperti per la prima volta in un determinato ospedale nell'arco di 2 anni.

Sono molte le ragioni che di solito impediscono l'inclusione di tutti gli individui che rispondono ai criteri diagnostici, di tempo e di luogo specificati. Alcune persone cambiano residenza, o muoiono, altri ancora non vogliono collaborare; oppure possono andare perdute alcune cartelle cliniche, cosicché informazioni importanti non sono più accessibili al ricercatore. Quest'ultimo dovrebbe pertanto riportare il numero dei casi selezionati in base ai criteri iniziali e il numero effettivo di persone reclutate e, inoltre, dovrebbe specificare i motivi delle omissioni e il numero dei casi omessi per ciascun motivo.

Se possibile, uno studio dovrebbe includere soltanto i casi incidenti, ossia quelli che si sviluppano o vengono diagnosticati per la prima volta durante il periodo di raccolta dei dati. Si eviterebbe così la mancanza di rappresentatività dei casi prevalenti. Utilizzando soltanto casi nuovi e scegliendo controlli che siano rappresentativi di soggetti a rischio di sviluppare la malattia, lo studio caso-controllo mira a identificare i fattori responsabili dello sviluppo della malattia stessa, proprio come uno studio di coorte. Talvolta, purtroppo, i casi incidenti, pur essendo più appropriati ai fini dello studio, sono in numero troppo esiguo per poter fornire risultati affidabili, soprattutto se la malattia è molto rara. Inoltre, per alcune malattie croniche come l'ipertensione, vi possono essere difficoltà insormontabili a individuare un periodo effettivo di insorgenza, il che rende necessario, ancora una volta, studiare i casi prevalenti. Si dovrà prestare allora particolare attenzione a determinare se un fattore associato alla malattia prevalente ne aumenti la durata anziché (od oltre a) causarne lo sviluppo o precederlo.

La scelta dei controlli ovvero del gruppo o dei gruppi di controllo rappresenta la difficoltà maggiore in questo tipo di studio e richiede un'ottima capacità di valutazione e una grande esperienza. In uno studio trasversale o di

coorte questo problema non si pone perché i casi vengono confrontati con tutti i soggetti non affetti presenti nella popolazione.

Quando decide per uno studio caso-controllo, più semplice e meno costoso, il ricercatore spera di poter ottenere ugualmente informazioni sulla relazione tra la malattia e le altre variabili confrontando un gruppo di casi e uno di controlli. Idealmente i controlli dovrebbero essere un campione della popolazione da cui provengono i casi, ossia essere rappresentativi delle persone che sarebbero state incluse nel gruppo dei casi se avessero sviluppato la malattia in oggetto. Qualora questa premessa non sia realizzabile, si cerca di avvicinarsi a essa il più possibile.

Una delle considerazioni più importanti circa la scelta dei controlli riguarda le informazioni che devono essere raccolte sulle variabili da studiare o sui possibili fattori eziologici.

Per quanto riguarda la qualità e la disponibilità di queste informazioni non dovrebbero esserci grosse differenze fra il gruppo dei casi e quello dei controlli. Disponibilità delle informazioni significa poter conoscere:

- 1) la quantità di informazioni raccolte per ogni caso e per ogni controllo;
- 2) quale percentuale nel gruppo dei casi e dei controlli può fornire tali informazioni.

Analoghe possibilità di accedere a importanti informazioni registrate con modalità simili per i casi e per i controlli (per esempio, la data di nascita o il peso registrati da uno stesso ospedale) può favorire la scelta di un particolare gruppo di controlli. Se le informazioni si ottengono attraverso interviste potrebbero esserci delle differenze fra i casi e i controlli, dovute alle condizioni emotive, alla conoscenza della malattia oggetto di studio, al livello di istruzione, alle condizioni socio-economiche e al luogo in cui l'intervista viene condotta (per esempio in casa o in ospedale).

Il ricercatore deve scegliere controlli il più possibile simili ai casi, eccetto per l'assenza della malattia presa in esame, tenendo conto di possibili fonti conosciute di vizi circa la qualità e quantità delle informazioni relative ai casi e ai controlli e tenendo in considerazione anche i vizi non noti. Tuttavia lo sforzo di cercare controlli molto simili ai casi non deve essere spinto fino al punto in cui diventa quasi impossibile differenziare i primi dai secondi per quanto riguarda i fattori che vengono studiati. Per esempio, scegliendo controlli che abbiano lo stesso livello di istruzione dei casi si minimizzano le differenze nella comprensione del questionario scritto, ma viene preclusa la possibilità di studiare il rapporto tra il livello di istruzione e la malattia e si può anche alterare il confronto caso-controllo per quanto riguarda fattori legati all'istruzione, come per esempio le condizioni socio-economiche. Quando i controlli vengono selezionati in maniera tale da risultare talmente simili ai casi per una caratteristica da far sì che ciò mascheri la relazione fra la malattia e un'altra caratteristica, si dice che i casi e i controlli sono *sovra-appaiati*. Ciò può accadere quando i controlli vengono appaiati ai casi per una caratteristica che è connessa con la possibile causa della malattia studiata, ma è una causa non indipendente della malattia o fa parte del percorso attraverso cui la possibile causa presa in esame porta alla malattia stessa.

Supponiamo che un basso livello socio-economico porti a una malattia attraverso un ridotto consumo di frutta fresca. Nell'esempio sopra citato, un appaiamento in base al livello di istruzione costituisce un sovra-appaiamento, in quanto un basso livello di istruzione, correlato a un basso livello socio-economico, non porta di per sé alla malattia. Anche se appaiassimo i due gruppi in base al consumo di frutta fresca, il nostro studio della relazione tra livello socio-economico e malattia sarebbe sovra-appaiato e la relazione stessa verrebbe mascherata, in quanto il consumo di frutta fresca si trova lungo la catena causale tra livello socio-economico e malattia. Si parla di sovra-appaiamento anche quando l'appaiamento riduce l'attendibilità statistica o aumenta i costi di uno studio.

Nella selezione del gruppo di controllo bisogna rispondere a due importanti domande:

- 1 Da quali fonti devono essere tratti i controlli?
2. Quale deve essere il metodo di selezione dei controlli da ogni fonte?

Queste decisioni devono tenere conto della necessità cui abbiamo accennato prima, di una certa similitudine tra i controlli e i casi, e di altre considerazioni pratiche, come la potenziale disponibilità dei controlli e le risorse, sia umane che finanziarie, disponibili per lo studio.

La scelta della fonte da cui ottenere i controlli:

1. Pazienti appartenenti alla stessa unità sanitaria:

- a) indipendentemente dalla diagnosi;
- b) escludendo quelli con particolari malattie;
- c) includendo solo quelli affetti da malattie che si pensa non siano associate ai fattori causali che vengono esaminati;
- d) coloro che risultano sani a una visita medica.

2. Persone reclutate al di fuori di strutture sanitarie:

- a) un campione della popolazione generale;
- b) soggetti che lavorano nelle organizzazioni sanitarie che segnalano i casi;
- c) amici e conoscenti;
- d) colleghi;
- e) vicini di casa;
- f) membri della famiglia, mogli o fratelli.

Per quanto al momento della scelta di un campione si prendano in considerazione i pro e i contro di ogni fonte è solo a studio concluso che, attraverso le critiche, possono emergere le ragioni per cui la scelta di un particolare gruppo ha viziato i risultati. Per esempio, il ricercatore decide di scegliere i controlli per i casi di calcolosi renale ospedalizzati fra i pazienti operati di plastica erniaria nello stesso ospedale, perché a quell'ospedale fa riferimento un particolare gruppo etnico e socio-economico della comunità, e perché dopo che il dolore acuto si è attenuato, le condizioni psicologiche di un paziente affetto da calcolosi renale non sono molto diverse da quelle di un paziente operato per ernia (altrettanto non si può dire per un paziente colpito da ictus o un malato terminale di cancro). Se viene riscontrata un'importante differenza fra i pazienti affetti da calcolosi renale e quelli affetti da ernia bisogna domandarsi se la differenza sia dovuta alla calcolosi renale o all'ernia. Per questo sarebbe opportuno disporre di un gruppo di controllo eterogeneo dal punto di vista della diagnosi o, se possibile, di più di un gruppo. Analogamente, la ripetizione del medesimo studio da parte di altri ricercatori, in condizioni diverse, può confermare se i risultati ottenuti sono validi.

La selezione dei controlli dalla fonte prevede un campionamento. Se le risorse sono limitate, il gruppo dei controlli sarà di dimensioni paragonabili al gruppo dei casi o, se è necessario, più piccolo. Se le risorse permettono di studiare un maggior numero di individui, ma non sono disponibili altri casi, il gruppo dei controlli può essere ampliato fino a raggiungere, per esempio, una dimensione pari al doppio o al triplo del gruppo dei casi. In tal caso si riduce notevolmente la variabilità campionaria. Sostanziali riduzioni della variabilità campionaria si possono ottenere ampliando il gruppo dei controlli fino a circa 5 o 6 volte il gruppo dei casi. Aumentando ulteriormente il gruppo dei controlli non si ottengono sensibili aumenti dell'attendibilità. Come già osservato precedentemente la scelta di un'unica fonte pone dei limiti di ordine generale alla natura del gruppo dei controlli. Inoltre, scegliendo i controlli dalla fonte, il ricercatore li appaierà con i casi tenendo conto di importanti caratteristiche quali l'età e il sesso. Appaiando casi e controlli in ordine a determinate caratteristiche, il ricercatore cerca di eliminare la possibilità che una differenza caso-controllo per questa caratteristica possa contribuire alla differenza caso-controllo per la variabile in esame. Sebbene questo approccio sia intuitivamente interessante, non sempre esso raggiunge il proprio scopo negli studi caso-controllo.

I controlli vengono scelti uno a uno per essere poi appaiati ai casi. Ovvero, per ogni caso vengono selezionati uno o più controlli con metodi sistematici seguendo precise regole e criteri. In uno studio sulla calcolosi renale, si può decidere di scegliere come controlli altri pazienti urologici che non abbiano calcoli delle vie urinarie o disturbi mentali dovuti a uremia o a tre cause, e che vengano appaiati con i casi in base all'età, al sesso, alla razza e alla data di ricovero. L'appaiamento individuale di un controllo per ciascun caso può essere eseguito scegliendo il primo paziente ammesso al servizio urologico dopo il caso che soddisfi i criteri richiesti per quanto concerne la diagnosi e le condizioni psichiche, che sia della stessa razza e dello stesso sesso e la cui età non differisca di più di 5 anni rispetto al caso. Quando si effettua un appaiamento per variabili quantitative, come l'età o la data di ricovero, è ammissibile una certa elasticità, altrimenti si rischia di non avere controlli per la maggior parte dei casi. Lo stesso può accadere se si effettua un appaiamento che tiene conto di molte caratteristiche.

Un'altra forma di appaiamento talvolta utilizzata è nota come *appaiamento di frequenza*. Qui il ricercatore non cerca di ottenere un appaiamento individuale dei controlli ai casi, ma seleziona i controlli che presentano una particolare caratteristica in modo da mantenere il loro numero uguale o proporzionale al numero dei casi con quella data caratteristica. Pertanto, se il gruppo dei casi è composto da 20 persone di età compresa tra i 40 e i 50 anni, 40 persone di età compresa tra i 50 e i 60 anni e 70 persone di età compresa tra i 60 e i 70 anni, il gruppo dei controlli può essere selezionato in modo da avere 20, 40 o 70 persone, o eventualmente 40, 80 e 140 persone (il doppio) nelle rispettive decadi di età. Come l'appaiamento individuale, quello di frequenza contribuisce ad assicurare che i casi possano essere confrontati, per ciascun livello della variabile di appaiamento, con un numero di controlli sufficiente a garantire l'attendibilità statistica dei risultati. Se vi fossero 70 casi e, per esempio, solo 3 controlli di età compresa tra i 60 e i 70 anni, il confronto in questo gruppo di età non potrebbe essere ritenuto molto attendibile.

Un metodo divenuto d'uso comune per identificare controlli nella popolazione generale è la selezione randomizzata di numeri telefonici. Si compongono numeri telefonici della comunità dove si sono manifestati i casi: spesso si seleziona dapprima in maniera randomizzata una serie di 100 numeri con lo stesso prefisso e le prime 5 cifre, scegliendo quindi in maniera casuale le ultime due cifre al fine di chiamare prima un numero e successivamente altri numeri della serie qualora il primo non risultasse corrispondere a un'abitazione. Se nei nuclei familiari così raggiunti sono presenti dei controlli adatti, essi vengono contattati e, se disposti a collaborare, sottoposti a un colloquio o a un test in maniera analoga ai casi. Questa procedura non è semplice come sembra, e comporta spese e fatica.

Se la malattia studiata è poco frequente nel gruppo che serve come fonte dei controlli, per escluderla può bastare una documentazione minima. I questionari per la raccolta delle informazioni dovrebbero essere strutturati in modo tale da evitare riferimenti alla malattia, almeno fin quando non siano state formulate le domande sulle variabili eziologiche.

Per la raccolta dei dati può essere utilizzata qualsiasi fonte di informazione per le variabili che si vogliono studiare. L'ideale sarebbe disporre di informazioni sui casi e sui controlli raccolte prima dello sviluppo della malattia. A volte può rendersi necessario raccogliere le informazioni dopo la manifestazione della malattia, ma dovrebbe essere fatto ogni sforzo per evitare differenze sia qualitative che quantitative fra i casi e i controlli nella acquisizione dei dati. Per esempio, se fosse possibile, gli assistenti che raccolgono gli esiti degli esami di laboratorio non dovrebbero sapere se le informazioni si riferiscono ai casi o ai controlli; analogamente, i questionari per la raccolta delle informazioni dovrebbero essere strutturati in modo tale da evitare riferimenti alla malattia, almeno fin quando non siano state formulate le domande sulle variabili eziologiche.

L'analisi dei dati sono espressi in termini di proporzione di casi e proporzione di controlli che hanno una particolare caratteristica. Se la caratteristica esaminata è quantitativa piuttosto che qualitativa (presenza o assenza di un attributo), la distribuzione fra i casi e i controlli può essere confrontata attraverso le proprietà delle distribuzioni: medie, deviazioni standard, mediane.

Considerando una determinata caratteristica come possibile fattore eziologico di una malattia, è spesso più significativo o naturale esprimere l'associazione della caratteristica con la malattia nello stesso modo in cui viene espressa in uno studio di coorte, cioè, come confronto della percentuale di soggetti malati fra gli individui con e senza quella caratteristica. Sfortunatamente i tassi di incidenza non sono direttamente ottenibili dagli studi caso-controllo. Tuttavia, mediante semplici formule è possibile calcolare dai dati di uno studio caso-controllo i tassi di confronto, cioè sia il rischio relativo che la frazione attribuibile.

Normalmente, gli studi caso-controllo valutano i rischi relativi e i rapporti tra tassi utilizzando l'odds ratio (rapporto delle probabilità, o rapporto crociato), che li approssima discretamente quando la malattia è relativamente rara. Ciò vale per molte malattie croniche che compaiono in meno del 5% delle popolazioni normali nell'arco di diversi anni di osservazione. In realtà, si selezionano controlli a rischio al momento in cui viene diagnosticato ciascun caso, si possono stimare direttamente i rapporti tra tassi istantanei. Se il campione dei controlli viene selezionato da tutti i soggetti della popolazione di provenienza dei casi, inclusi quelli che sono divenuti casi, si possono calcolare direttamente i rapporti di rischio cumulativo. Ciò nonostante, gli odds ratio vengono impiegati di frequente, e dovrebbero essere noti a tutti gli operatori sanitari.

Il metodo del rapporto crociato per calcolare il rischio relativo da uno studio caso-controllo è indipendente da un precedente appaiamento fra casi e controlli. L'odds ratio, o relative odds si può determinare attraverso la seguente formula:

$$\text{Odds ratio} = \frac{\text{numero di casi con una caratteristica}}{\text{numero di casi senza quella caratteristica}} \times \frac{\text{numero di controlli senza quella caratteristica}}{\text{numero di controlli con la caratteristica}}$$

Nella tabella sotto i numeri sono rappresentati dalle lettere a, b, c, d; in questo caso il rapporto crociato è $a*d / b*c$. I numeri fra parentesi, sempre nella tabella sotto, si riferiscono a un esempio specifico che comprende 100 casi e 100 controlli. La caratteristica è presente in 20 dei casi e in 10 dei controlli; quindi il rapporto crociato è $(20 * 90)/(80 * 10) = 2,25$. Da queste informazioni possiamo stimare che le persone con quella data caratteristica hanno una probabilità di contrarre la malattia 2,25 volte superiore rispetto a coloro che non hanno quella caratteristica.

Tabella: Risultati di uno studio caso-controllo

Caratteristica	Casi	Controlli
Presente	a (20)	c (10)
Assente	b (80)	d (90)
Totale	100	100

Interpretazione e accorgimenti.

Se tra i casi una caratteristica viene riscontrata con maggior frequenza che nei controlli (cioè il rischio relativo è maggiore di 1), o se le distribuzioni o i livelli medi di una data caratteristica differiscono, allora esiste un'associazione fra la caratteristica in esame e la malattia.

E' bene sottolineare che la natura dei casi studiati è influenzata anche dalla loro origine. Per esempio, i casi reclutati solo tra i ricoverati di un ospedale possono essere affetti dalla malattia in forma più grave dei pazienti ambulatoriali. Quindi, come abbiamo sottolineato per quanto riguarda il gruppo dei controlli, anche le caratteristiche del gruppo dei casi devono essere considerate con la massima attenzione nella progettazione dello studio e nell'interpretazione dei risultati.

Sebbene l'appaiamento sia un metodo di controllo per le variabili estranee o di confondimento apparentemente semplice, esso può introdurre complicazioni inaspettate e errori in uno studio caso-controllo. Pertanto non dovrebbe essere effettuato se non fornisce un contributo importante, ma nel caso in cui venga effettuato dovrà essere eseguita l'analisi per coppie o per unità di appaiamento.

Valutazione e ruolo del metodo caso-controllo.

Gli studi caso-controllo sono generalmente i più veloci e meno costosi tra gli studi epidemiologici analitici. Rappresentano l'unico approccio pratico per malattie rare. Tuttavia problemi che insorgono nel reperire un gruppo di casi rappresentativo e un appropriato gruppo di controlli e nel raccogliere informazioni paragonabili sui casi e sui controlli, sono spesso così complessi che sui risultati di uno studio caso-controllo si possono sollevare molte e legittime obiezioni, più che per gli studi trasversali o per quelli di coorte.

Gli studi caso-controllo hanno svolto un ruolo primario nello sviluppo di molte e fruttuose ricerche. Per esempio, l'associazione fra fumo di sigaretta e cancro del polmone venne dimostrata con uno studio caso-controllo, prima che venissero intrapresi studi di coorte. Dato il loro basso costo rappresentano, in molte circostanze, il primo approccio per verificare alcune ipotesi. Sono anche molto usati nelle indagini esplorative con molte variabili (qualche volta vengono infatti chiamati fishing expedition) per trovare indizi e orientamenti per studi futuri.

Un'altra applicazione è il caso-controllo nidificato (nested). In uno studio di coorte molto grande, per le difficoltà, la spesa, o la non disponibilità di informazioni originarie, potrebbe essere più conveniente raccogliere i dati necessari solo per i casi che sviluppano la malattia e solo per un sottogruppo fra coloro che non la contraggono.

Esempio di indagine caso-controllo

Verifica del fattore di rischio rappresentato dal consumo di alcool per l'insorgenza della cirrosi epatica in una popolazione di 10.000.000 abitanti. I casi patologici risultano 25.000 (con una prevalenza di 0,25%), 20.000 dei quali sono consumatori abituali di alcool, mentre tra gli abitanti non cirrotici i bevitori sono 975.000.

Si organizza una tabella a due entrate:

	Casi (malati)	Controlli (non malati)	Totale
Esposti (bevitori)	20.000	975.000	995.000
Non esposti (non bevitori)	5.000	9.000.000	9.005.000
Totale	25.000	9.975.999	10.000.000

che permette di calcolare

$$\text{Prevalenza (esposti)} = \frac{20.000}{995.000} = 0.020$$

$$\text{Prevalenza (non esposti)} = \frac{5.000}{9.005.000} = 0.00056$$

Il rapporto tra le prevalenze porta al rischio relativo RR che indica quanto è più frequente la cirrosi tra chi era esposto rispetto a chi non lo è.

$$\frac{0.020}{0.00056} = 35.7$$

Questo significa che, il consumo di alcool ha un rischio di 36 volte di contrarre la cirrosi rispetto a coloro che non lo consumano. Se si applica la formula approssimata, il valore è lievemente diverso, anche se la sostanza del risultato non cambia:

$$RR = a/b * d/c = \frac{20*9.000}{975*5} = 36.9$$

Risulta del tutto evidente la difficoltà di studiare una popolazione così consistente, per cui si ricorrerà ad un campione: si organizzerà una tabella con le frequenze campionarie e si effettueranno i calcoli come in precedenza. Il calcolo rappresenterà una stima del rischio relativo.

c) Studi longitudinali (coorti)

Fra i vari tipi di studi epidemiologici osservazionali gli studi di coorte, o di incidenza, permettono di ottenere informazioni definitive sull'eziologia delle malattie, e misure dirette del rischio di sviluppare le malattie. Se vengono condotti in modo prospettico, possono essere costosi e molto lunghi, perché richiedono un impiego a lungo termine di fondi e di personale specializzato. Inoltre, come si dirà più avanti, essi non sono scevri da vizi potenziali e da altri problemi di carattere scientifico.

Come vengono condotti gli studi di coorte

In primo luogo occorre definire una popolazione, o coorte, da studiare. Al fine di osservare l'incidenza di una determinata malattia (o malattie) la popolazione viene seguita per un determinato periodo di tempo. In uno studio di coorte prospettico, la popolazione viene definita e caratterizzata in base alla sua situazione all'inizio dello studio, e viene seguita quindi per un certo periodo. In uno studio di coorte retrospettivo o storico, la popolazione viene definita e caratterizzata secondo la sua situazione nel passato, in base a dati già registrati, seguendola quindi verso il presente sino a un momento determinato. La coorte può essere composta da un gruppo eterogeneo di persone, per esempio una comunità, o da gruppi specifici più facilmente studiabili, per esempio persone che svolgono la stessa attività o che hanno stipulato un'assicurazione. La coorte può anche essere scelta sulla base dell'esposizione a un fattore eziologico sospetto, come le radiazioni ionizzanti, un farmaco. Se l'esposizione al fattore sospetto è una caratteristica comune a tutti o a quasi tutti i membri della

coorte, sarà necessario selezionare una seconda coorte simile di persone non esposte a quel fattore, od ottenere un altro standard di confronto per valutare l'esperienza del gruppo degli esposti.

Gli studi di coorte focalizzano l'attenzione sulla comparsa della malattia. Per osservare l'insorgenza di una malattia, questa deve, ovviamente, essere assente nella popolazione studiata all'inizio della ricerca.

E' necessario, pertanto, dimostrare in qualche modo che essa sia costituita da individui non malati ma a rischio di sviluppare la malattia. Nel caso di una malattia rara con decorso rapido e mortale, per esempio la leucemia acuta, è possibile individuare facilmente i pochi casi presenti nella popolazione. Per quanto riguarda, invece, malattie più comuni, per esempio la cardiopatia coronarica negli uomini di mezza età, è bene esaminare la popolazione potenzialmente oggetto dello studio per individuare, e quindi escludere, coloro che sono già affetti dalla malattia di cui si vuole documentare l'incidenza. Tuttavia alcuni studi di coorte con obiettivi specifici possono non richiedere esami iniziali in quanto le informazioni necessarie per caratterizzare i soggetti studiati provengono da altre fonti.

L'osservazione degli individui nel tempo, o 'follow-up'

Una volta che la popolazione sia stata definita e siano state valutate le caratteristiche dei soggetti, essa deve essere seguita per osservare la comparsa della malattia. La modalità con cui viene condotto il follow-up variano da studio a studio, sia quanto a frequenza che a completezza, in rapporto alle manifestazioni della malattia che si vogliono valutare.

Follow-up semplici, ma relativamente completi, vengono condotti dalle compagnie di assicurazione sulla vita, per indagare sui fattori che influenzano la mortalità. Ai loro fini la morte è l'unico punto di arrivo importante e deve essere notificata alla compagnia per il pagamento della somma stabilita dalla polizza.

Se invece si vogliono rilevare tutti i nuovi casi di cardiopatia coronarica o di ictus, sono necessarie molte altre procedure, come esami periodici ripetuti, il controllo di decessi, ricoveri e visite mediche, e bisogna mantenere i contatti con coloro che si trasferiscono altrove. Tuttavia, il fatto di avere a disposizione risorse limitate può far sì che solo alcune di queste procedure di controllo siano realizzabili, per esempio solo per i ricoveri e le morti. Anche se incompleti, questi follow-up parziali possono essere perfettamente adeguati agli scopi dello studio.

La durata del follow-up viene stabilita sulla base del numero di casi che sono necessari per rispondere in modo attendibile e statisticamente significativo ai quesiti specifici cui la ricerca vuole rispondere. Può essere stabilita inizialmente, se si conosce la grandezza della popolazione e il tasso di incidenza della malattia. Per esempio, se la popolazione studiata è di 1000 persone e il tasso di incidenza è dell'1% ogni anno, dobbiamo aspettarci di individuare 10 nuovi casi per ogni anno di follow-up. Se sono necessari 100 casi per ottenere delle risposte con un certo grado di attendibilità, lo studio dovrà essere protratto per 10 anni.

Questo esempio è una semplificazione e non tiene conto di una possibile riduzione del numero di nuovi casi per anno dovuta ad abbandono da parte dei soggetti studiati, o di un possibile aumento, negli anni, dei nuovi casi per la malattia osservata, causato dal progressivo invecchiamento della popolazione, se l'incidenza aumenta con l'età. Sebbene dal punto di vista pratico sia consigliabile che un follow-up duri il minor tempo possibile, uno studio può richiedere un periodo di osservazione molto lungo per determinare i fattori che provocano o predicono la malattia in un futuro non immediato. Ciò può essere necessario per quelle malattie, tra cui molte forme di cancro, che presentano un lungo periodo di induzione tra l'esposizione a un fattore causale e il conseguente aumento di incidenza. Il periodo di induzione tra l'esposizione e l'insorgenza della malattia viene comunemente detto latenza, sebbene tale termine si riferisca più precisamente a un periodo in cui la malattia, pur essendo già presente, è asintomatica o non viene diagnosticata.

Durante il periodo di follow-up è possibile ripetere le valutazioni iniziali sulle caratteristiche della popolazione. In questo modo la comparsa della malattia può essere studiata sia in rapporto alle caratteristiche iniziali che a possibili modificazioni di queste. Per esempio, può essere interessante dimostrare non solo se il livello di colesterolo nel siero sia associato alla successiva comparsa di cardiopatia coronarica, ma anche se un aumento o una diminuzione di questo possa essere in qualche modo maggiormente predittivo.

Vi sono anche altre ragioni per riesaminare le caratteristiche di una popolazione durante il periodo di osservazione. Nel corso di un follow-up molto lungo possono essere apportati dei miglioramenti alle tecniche di misurazione, o nuove conoscenze sulla malattia possono suggerire l'opportunità di misurare variabili aggiuntive non incluse inizialmente nello studio.

Analisi dei dati.

Negli studi di coorte, come in uno studio trasversale, la popolazione è suddivisa e classificata secondo le variabili che si pensano essere legate alla malattia. Si determina il tasso di incidenza per ogni sottogruppo, e i tassi vengono confrontati per verificare se la presenza o l'assenza della variabile (o la differenza nei livelli, se la variabile è quantitativa) sia da porre in relazione alla successiva comparsa della malattia. Se la popolazione in esame è una coorte esposta a un fattore eziologico sospetto, l'incidenza della malattia deve essere confrontata con quella di una coorte simile di non esposti, o della popolazione generale.

Se quasi tutti, o tutti, i membri di una popolazione vengono seguiti per lo stesso periodo di tempo, allora può essere utilizzato un semplice tasso di incidenza cumulativa. Per esempio, se il periodo di follow-up è di 3 anni per tutti gli individui, viene calcolato il tasso di incidenza relativo a 3 anni in ogni sottogruppo. Se vi sono sostanziali differenze nella durata del follow-up fra i soggetti in studio, queste dovranno essere prese in considerazione nell'analisi dei dati. La durata di un follow-up può essere notevolmente diversa quando i soggetti abbandonano lo studio prima che questo sia concluso, per esempio, se essi cambiano residenza o muoiono. In alcune ricerche è necessario aggiungere nuovi soggetti alla popolazione in studio per un periodo relativamente lungo di osservazione (ciò porta a una cosiddetta coorte aperta: questa si differenzia dalla coorte chiusa in cui il gruppo da studiare, stabilito in modo definitivo, viene determinato all'inizio dell'osservazione). Ciò significa che se l'incidenza della malattia viene determinata in un preciso momento, i soggetti saranno stati seguiti per periodi diversi a partire dal loro ingresso nello studio.

Il metodo standard, quando si lavora su periodi di follow-up diversi, è quello di usare, come denominatore del tasso di incidenza, gli anni-persona di osservazione (mesi-persona, o giorni persona ecc.) se sono periodi più appropriati e convenienti. In questo modo, ogni soggetto contribuisce per il periodo di tempo in cui è stato sotto osservazione; se un individuo lascia lo studio dopo 1 anno, egli contribuisce per 1 anno-persona; se lo lascia dopo 10 anni, contribuirà per 10 anni-persona.

Una caratteristica apprezzabile del metodo degli anni-persona consiste nel fatto che un soggetto può contribuire a più di un sottogruppo con anni-persona di osservazione. Si supponga, per esempio, che in uno studio della durata di 5 anni si determini l'incidenza di una malattia per sottogruppi divisi in base alle decadi di età. Una persona entrando nella popolazione in studio all'età di 48 anni, contribuisce per 2 anni-persona di osservazione al sottogruppo degli individui con età compresa fra i 40 e i 49 anni, e per 3 anni-persona di osservazione al sottogruppo di individui che hanno fra i 50 e i 59 anni. Questo può verificarsi anche con altre misure che cambiano nel tempo. Un individuo può infatti rimanere alcuni anni in un particolare quartile di colesterolemia e passare nel corso del tempo in un quartile più alto o più basso.

Il tempo di follow-up deve essere limitato al periodo in cui i soggetti sono effettivamente a rischio.

E' ovvio che il periodo successivo alla morte non è più un periodo a rischio per il soggetto studiato.

E' tuttavia facile cadere nella trappola di includere erroneamente nello studio il periodo precedente a quello in cui il soggetto è a rischio.

Interpretazione e valutazione degli studi di coorte

Lo scopo degli studi di coorte è di migliorare la capacità di previsione della comparsa di una malattia. Questo tipo di ricerca dimostra chiaramente la sequenza temporale fra la presenza, o l'assenza, di una particolare caratteristica e la successiva comparsa della malattia. Comunque il prevedere la malattia non implica necessariamente un rapporto di causa-effetto. Come è già stato sottolineato, i fattori di rischio associati a una malattia per predirla debbono precederla, come negli studi trasversali o caso-controllo.

Un problema che è stato evidenziato a proposito degli studi trasversali e caso-controllo è la possibilità di una sovrastima di casi a decorso prolungato. Negli studi di coorte questo problema non si pone perché in tali studi gli individui esposti e non esposti vengono seguiti nel tempo.

Nonostante siano ampiamente apprezzati, gli studi di coorte sono soggetti a errori. Abbiamo già detto come negli studi trasversali o caso-controllo, la presenza o l'assenza di una malattia possano influire sul fattore studiato o sulla sua misurazione. In uno studio di coorte, viceversa, la presenza o l'assenza di un fattore studiato può influenzare la successiva valutazione della malattia.

In uno studio sull'ictus, per esempio, è possibile che il fatto di conoscere la pressione arteriosa degli individui condizioni, consciamente o inconsciamente, un'eventuale diagnosi di vasculopatia cerebrale da parte del ricercatore. Se questo si verifica, lo studio conterrà a priori la correlazione fra pressione arteriosa e incidenza di ictus. Analogamente, in uno studio sul cancro, poiché il fatto di diagnosticare la malattia dipende in qualche misura dall'iniziativa o dalla disponibilità dei soggetti a sottoporsi a visite mediche, gli individui con anamnesi familiare positiva per cancro o i fumatori saranno più motivati a sottoporsi a un controllo; ciò darà luogo a

errori, o a una correlazione precostituita fra cancro e precedenti di tale malattia in famiglia, o tra cancro e fumo. Deve pertanto essere fatto ogni sforzo affinché la diagnosi sia formulata in modo indipendente dai possibili fattori eziologici studiati.

Gli studi di coorte sono soggetti a possibili vizi per la perdita di soggetti compresi nello studio, sia per un rifiuto iniziale di parteciparvi, sia per la perdita di individui in studio durante il periodo d'osservazione. Non è detto che tali perdite invalidino lo studio. I ricercatori, però, dovrebbero considerare attentamente i motivi che portano alla perdita di individui e verificare che essi non interferiscano con i risultati dello studio. Qualche volta è possibile ottenere informazioni dall'esterno riguardo alla perdita dei soggetti, in particolare se il loro abbandono sia dovuto a malattia o morte o a motivi riconducibili alle variabili e alla malattia studiata.

Lo studio di Framingham

Quando lo studio di Framingham iniziò, intorno al 1950, questa città del Massachusetts aveva circa 28000 abitanti. C'erano molte ragioni per condurre uno studio di questo tipo proprio in quella città. In quel periodo, Framingham era una comunità quasi autosufficiente, con zone industriali e agricole. Per questa e altre ragioni non la si poteva considerare atipica. Vi era un numero sufficiente di residenti nella fascia di età utile a fornire un gruppo di individui adeguato allo studio. Si riteneva, da un precedente studio sulla tubercolosi condotto con risultati positivi e dalle dichiarazioni dei medici e delle persone interpellate a riguardo, che la popolazione avrebbe collaborato. La città era abbastanza piccola perché i residenti venissero sottoposti agli esami in un unico ambulatorio centrale. L'osservazione dei malati ricoverati era facilitata dal fatto che la maggior parte di essi confluiva nell'ospedale centrale della città. Inoltre Framingham dista solo una trentina di chilometri dai grandi centri medici di Boston, così da rendere possibile una eventuale consulenza a livello medico e scientifico. Nella programmazione dello studio fu previsto un periodo di osservazione di circa 20 anni, dato il lento instaurarsi dell'aterosclerosi e delle sue complicanze.

I limiti di età minimi e massimi della popolazione studiata furono fissati a 30 e 60 anni. Si ritenne opportuno escludere le persone più anziane poiché molte di loro avevano già un'aterosclerosi coronarica avanzata e lo studio avrebbe potuto rilevare solo eventuali fattori precipitanti. Le persone sotto i 30 anni vennero escluse perché l'incidenza di cardiopatie coronariche era molto bassa e la loro possibile mobilità avrebbe reso difficile il follow-up.

I ricercatori si proponevano di ottenere come campione un gruppo di 5000 individui; da questo numero di soggetti di età compresa tra i 30 e i 60 anni, nell'arco dei 20 anni, sarebbe emerso un numero di casi adeguato. Sapendo che qualcuno non avrebbe collaborato, i ricercatori selezionarono un campione sistematico più grande, comprendente i 2/3 dei 10000 abitanti appartenenti alla fascia d'età prefissata. L'elenco degli abitanti fu compilato in base alla circoscrizione elettorale e, all'interno di questa, in base alla dimensione del nucleo familiare (uno, due, tre o più membri di età compresa fra i 30 e i 60 anni). Venivano scelte due famiglie ogni tre. Era stato adottato il metodo di scegliere famiglie piuttosto che individui perché:

- 1) un individuo appartenente alla fascia d'età considerata non avrebbe rifiutato di sottoporsi agli esami se questi fossero stati proposti anche a un altro membro della sua famiglia;
- 2) molti uomini riluttanti potevano essere convinti dalle loro mogli a recarsi insieme all'ambulatorio;
- 3) veniva così favorito lo studio di particolari caratteristiche di coppie di coniugi e di gruppi familiari.

I 6507 individui che formavano il campione furono invitati a partecipare allo studio da concittadini incaricati di reclutare le persone che vivevano nel loro stesso quartiere. I reclutatori erano un gruppo di volontari che si era presentato per eseguire esami cardiovascolari prima dell'inizio ufficiale dello studio. Avendo già avuto un'esperienza personale essi potevano descrivere ad altri in che cosa consistessero gli esami e le visite mediche richieste dallo studio.

Nonostante l'approccio personale, solo 4469 individui, circa i 2/3 del campione, accettarono di partecipare. A essi si aggiunse un gruppo di 740 volontari, e il numero totale fu quindi di 5209 individui. Gli esami iniziali rilevarono che 82 soggetti erano già affetti da evidente cardiopatia coronarica; questi vennero esclusi dalla popolazione a rischio e il gruppo finale si ridusse a 5127 persone.

Tutti vennero sottoposti a esami completi ogni due anni. La valutazione comprendeva l'anamnesi, una visita medica, ed esami di laboratorio o strumentali pertinenti quali la misurazione dei lipidi nel siero, l'elettrocardiogramma e la radiografia del torace. Questi esami dovevano permettere di evidenziare lo sviluppo di cardiopatie coronariche e di altre malattie dovute all'aterosclerosi come ictus e le vasculopatie periferiche. Ogni due anni venivano controllate anche le variabili legate al manifestarsi della malattia. A mano a mano che i nuovi metodi di indagine acquisivano importanza in questo campo della ricerca, essi venivano aggiunti a quelli effettuati periodicamente. I ricercatori non si limitavano alle informazioni sulle possibili variabili eziologiche

ricavate nel primo esame. Vennero mantenuti i contatti con la popolazione in studio e con i medici della città. I tempi di attesa per l'esecuzione degli esami erano ridotti il più possibile e un rapporto completo sul loro esito veniva inviato al medico di fiducia dei soggetti in studio. I medici che eseguivano gli esami non prescrivevano alcuna terapia e non elargivano consigli di alcun genere, tranne l'invito a rivolgersi al medico di famiglia nei casi in cui fossero state riscontrate gravi anomalie.

Sebbene gli esami biennali costituissero la principale fonte di informazione, l'insorgenza della malattia fu rilevata anche in altri modi, come per esempio tramite le cartelle cliniche dei soggetti ricoverati, le visite presso gli ambulatori dei medici del luogo, i certificati di morte, gli esami necroscopici e le dichiarazioni dei familiari. La diagnosi per ogni malattia studiata veniva posta in base a criteri rigorosi, così da includere fra i malati solo i casi certi.

Lo studio ha confermato con grande evidenza che la malattia aterosclerotica non colpisce le persone anziane in modo casuale con il progredire degli anni, e che gli individui predisposti possono essere individuati prima che abbia luogo qualsiasi manifestazione clinica. Indicazioni sulla suscettibilità, o "fattori di rischio", individuati nello studio di Framingham e in altre indagini epidemiologiche, includono: il sesso maschile, l'età avanzata, un'elevata concentrazione di lipidi nel siero, l'ipertensione, il fumo, il diabete mellito (o anche intolleranze ai carboidrati di minore entità), l'obesità, la bassa capacità vitale, un'elevata concentrazione ematica di fattori della coagulazione e alcune alterazioni elettrocardiografiche. Altri fattori di rischio messi in evidenza da altri studi sono: fattori psico-sociali, anamnesi familiare positiva per cardiopatie coronariche e sedentarietà.

Le informazioni dettagliate ottenute sulla vasta popolazione di Framingham hanno permesso indagini più specifiche sul ruolo di ciascun fattore di rischio. Per esempio, fu scoperto che l'obesità non è associata nello stesso modo a tutte le manifestazioni della malattia coronarica; essa sembrerebbe predisporre all'angina pectoris e alla morte improvvisa, ma ha un legame più debole con l'infarto miocardico, in particolare nei maschi. Il numero di casi incluso nello studio è stato sufficientemente grande da permettere di analizzare anche le relazioni multiple tra diversi fattori di rischio. Un importante risultato è stato il poter mettere in evidenza che persone in cui sia presente più di un fattore di rischio (per esempio, uomini ipertesi che fumano e hanno elevate concentrazioni di lipidi nel sangue) sono esposte a un rischio particolarmente elevato di sviluppare malattie coronariche.

Con l'invecchiamento della popolazione fu possibile approfondire le conoscenze sulle malattie degli anziani, come ictus e la demenza senile. Inoltre, le informazioni raccolte durante lo studio di Framingham hanno permesso studi epidemiologici anche su malattie non aterosclerotiche, come il cancro, le cardiopatie reumatiche, la gotta, le colecistopatie, il glaucoma e la cataratta. Attualmente vengono studiati gli adulti nati dalla coorte originaria per valutare se le malattie cardiovascolari e i loro fattori di rischio siano legati a influenze familiari e genetiche e agli effetti dei cambiamenti, avvenuti nel tempo, nello stile di vita. Lo studio di Framingham ha inoltre permesso di condurre numerosi studi sui metodi epidemiologici.

Oggi gli studi epidemiologici sulle cardiopatie coronariche si stanno sempre più spostando da studi osservazionali, di cui quello di Framingham rappresenta uno dei più importanti, a studi sperimentali, che cercano il modo di ridurre il rischio di sviluppare la malattia.

Ruolo degli studi di coorte.

Dalla descrizione dello studio di Framingham dovrebbe risultare chiaro il perché studi prospettici di coorte sulle popolazioni generali siano poco frequenti. Essi sono di difficile esecuzione e molto costosi, richiedono la decisione iniziale di assumere un incarico a lunga scadenza e grande perseveranza sia da parte dei committenti che del personale coinvolto nello studio. Di solito, però, l'investimento si rivela proficuo per la qualità e la quantità delle informazioni che ne possono derivare. Si noti che gli studi di coorte consentono di studiare molte conseguenze relative alle esposizioni e ad altre caratteristiche di riferimento. Uno studio caso-controllo parte da casi che presentano una sola patologia o conseguenza finale ed esamina esclusivamente tale singola conseguenza.

La necessità di seguire una coorte per un lungo periodo o di osservare un gran numero di persone è resa più evidente dal fatto che molte malattie studiate in questo modo hanno mostrato tassi di incidenza sorprendentemente bassi. Le cardiopatie coronariche sono la maggiore causa di morte negli Stati Uniti, e l'aterosclerosi delle coronarie è un riscontro autoptico frequente negli uomini di mezza età, tuttavia l'incidenza di nuovi casi clinicamente identificati di cardiopatie coronariche fra uomini di mezza età è soltanto l'1% ogni anno. Così, sebbene l'ipertensione sia una malattia ad alta prevalenza fra gli adulti negli Stati Uniti, molti ipertesi sembrano divenire tali in modo graduale, ed è difficile quindi definire e identificare i nuovi casi nella popolazione in uno studio di coorte.

Gli studi di coorte retrospettivi possono essere condotti più rapidamente, se si identifica una coorte idonea della quale siano disponibili informazioni attendibili. Questo approccio è spesso usato negli studi epidemiologici sulla valutazione dei rischi occupazionali. Ciò nonostante molte malattie sono talmente rare che gli studi caso-controllo rappresentano attualmente l'unico approccio epidemiologico per indagare sulla loro eziologia. Probabilmente l'evoluzione tecnologica renderà gli studi di coorte più accessibili in futuro. L'archiviazione di informazioni mediche e demografiche nelle "banche dati" di supporti informatici sta diventando un metodo d'uso comune per aumentare l'efficienza e la qualità dell'assistenza medica. Ciò renderà disponibili molte informazioni su coorti che potranno essere studiate sia retrospettivamente che in modo prospettico. Il proseguimento degli studi nell'ambito del record linkage (cioè la combinazione di più informazioni per ogni individuo, come la data di nascita, le visite mediche, le prescrizioni di farmaci, le malattie e dati relativi alla morte) aumenteranno il numero delle diverse relazioni che possono essere studiate cioè delle relazioni tra le varie caratteristiche iniziali e i vari esiti delle malattie.

Esempio di uno studio di coorte

Viene presa in considerazione una coorte di 5.000 persone, delle quali 2.000 sono fumatori e 3000 non lo sono. La coorte viene seguita per un decennio, allo scopo di conoscere il tasso di incidenza del carcinoma polmonare in relazione all'esposizione al fumo di sigarette come fattore di rischio. Nel periodo considerato sviluppano un carcinoma polmonare 120 soggetti di cui 100 fumatori e 20 no.

	Cancro	Controlli	Totale
Fumatori	100	1.900	2.000
Non fumatori	20	2.980	3.000
Totale	120	4.880	5.000

I tassi di incidenza e i rischi ad essi collegati risultano:

$$\text{per i fumatori } I_{esp} = 100 / 2.000 * 1.000 = 50\text{‰}$$

$$\text{per i non fumatori } I_{non-esp} = 20 / 3.000 * 1.000 = 6.7\text{‰}$$

$$\text{RR(Rischio Relativo)} = \frac{I_{esp}}{I_{non-esp}} = \frac{50\text{‰}}{6.7\text{‰}} = 7.46$$

$$\text{RRA(Riduzione del Rischio Assoluto)} = I_{esp} - I_{non-esp} = (50 - 6.7)\text{‰} = 43.4\text{‰}$$

$$\text{RRR(Riduzione Rischio Relativo)} = \frac{I_{esp} - I_{non-esp}}{I_{esp}} = \frac{43.4\text{‰}}{50\text{‰}} = 0.87$$

Da questo studio di coorte deriva che:

- i fumatori mostrano un rischio di oltre 7 volte maggiore rispetto ai non fumatori;
- ci si aspetta una riduzione di circa 43 casi di tumore ogni mille fumatori in meno;
- eliminando il fumo si può ipotizzare una riduzione di circa l'87% dell'incidenza osservata tra i fumatori.

Studi sperimentali

Gli studi sperimentali assomigliano agli studi coorte per il fatto che anch'essi richiedono un follow-up dei soggetti studiati per documentare il verificarsi di alcuni eventi relativi al loro stato di salute. La loro caratteristica principale è tuttavia il fatto di implicare un intervento o una manipolazione da parte del

ricercatore, contrariamente agli studi di coorte e ad altri tipi di studi osservazionali, in cui il ricercatore non interviene ma si limita a osservare.

Si ritiene che gli studi sperimentali rappresentino il metodo migliore per valutare una relazione causa-effetto. Se dopo un certo intervento su di un gruppo sperimentale si osserva un determinato risultato, si presume che questo sia dovuto all'intervento stesso, a meno che la stessa cosa non capiti in un gruppo di controllo sul quale invece l'intervento non sia stato effettuato. La relazione causa-effetto può anche essere dimostrata eliminando, o riducendo, il fattore ritenuto causale in un gruppo sperimentale e mostrando la scomparsa o la riduzione dell'effetto, scomparsa o riduzione che non si verifica nel gruppo di controllo.

Se un fattore viene eliminato, o ridotto, e di conseguenza l'incidenza della malattia diminuisce, significa che in pratica quel fattore è una causa della malattia.

Sebbene venga attribuito grande valore ai loro risultati, condurre studi sperimentali può essere spesso estremamente difficile; essi inoltre sollevano spesso problemi etici che devono essere considerati con particolare attenzione.

Gli studi epidemiologici sperimentali riguardanti la medicina preventiva vengono definiti, da molti ricercatori, studi di campo (field trials) se gli interventi vengono effettuati su ogni singolo individuo del gruppo (es. le vaccinazioni) e studi comunitari (community trials) se gli interventi riguardano l'ambiente e le matrici biologiche (per esempio la fluorazione delle acque). Gli studi sperimentali riguardanti le sperimentazioni terapeutiche, l'uso di farmaci o l'esecuzione di interventi chirurgici, sono invece generalmente definiti studi clinici (clinical trials).

I problemi di carattere etico

Negli studi osservazionali il principale problema etico per il ricercatore, a parte l'oggettività e il lavoro scrupoloso necessari, è la riservatezza cui è tenuto circa i dati relativi a ciascun soggetto studiato. Sarebbe dannoso per un individuo che informazioni confidenziali raccolte per scopi medici o scientifici fossero accessibili ad altri, o che esse venissero comunicate all'interessato in modo improprio. In generale però, coloro che effettuano studi di questo tipo sono osservatori passivi della natura; per questa ragione le loro ricerche presentano pochi problemi etici.

La posizione etica degli sperimentatori è molto più complessa poiché essi intervengono in vari modi sulle persone, pertanto devono avere ragioni più che valide per ritenere che ciò che si propongono di fare abbia buone probabilità di comportare giovamenti. Contemporaneamente, essi devono porsi dei dubbi sul valore di quello che stanno per fare confrontandolo con ciò che veniva fatto in passato, oppure con il non fare nulla, altrimenti non potrebbero, in coscienza, sottoporre il gruppo dei controlli al trattamento tradizionale o a nessun trattamento.

Gli esperimenti medici possono essere condotti solo in una situazione di incertezza. Sfortunatamente, alcuni potenziali ricercatori sono così convinti dei benefici di una terapia o di una misura preventiva da non essere propensi a condurre un esperimento controllato per verificarne gli effetti. Queste sensazioni di certezza, spesso basate su dati sperimentali inadeguati, li portano a non voler privare il gruppo di controllo di tali trattamenti. Analogamente, alcuni medici irrazionalmente scettici, convinti della superiorità della terapia tradizionale o del non intervento, possono essere riluttanti a provare nuovi metodi su basi sperimentali. Queste persone dovrebbero convincersi del fatto che non effettuare un esperimento, quando se ne presenti la necessità e la possibilità, è altrettanto non etico.

La sensibilizzazione agli aspetti etici della sperimentazione sull'uomo ha portato alla formazione di comitati nelle università e in altre istituzioni di ricerca per controllare e approvare tutte le proposte di studi sperimentali sull'uomo. Attualmente vi è ampio accordo nel ritenere che, quando ciò sia possibile, gli individui debbano decidere in prima persona se partecipare allo studio. Questa decisione dovrebbe essere presa dopo aver valutato i rischi potenziali e i possibili benefici. Pertanto all'interessato, o ai parenti o tutori, viene richiesto il consenso informato.

Anche così, l'autonomia dei pazienti può essere compromessa se l'unica possibilità di ricevere il trattamento da essi preferito è quello di partecipare a uno studio in cui potrebbero venir assegnati in maniera randomizzata a un trattamento diverso.

Come viene condotto uno studio sperimentale

L'obiettivo principale dell'epidemiologia sperimentale è quello di valutare l'efficacia di misure utili a prevenire una malattia. Perché tale valutazione sia possibile, tali misure preventive devono essere applicate a un gruppo di individui. L'incidenza della malattia o delle conseguenze che essa comporta, come per esempio la disabilità,

viene misurata nel gruppo sperimentale, o gruppo trattato. Perché sia fonte di informazioni utili, un esperimento deve essere controllato; cioè il risultato deve essere confrontato con uno standard per determinare se il trattamento abbia comportato qualche beneficio. Lo standard può essere rappresentato da un altro gruppo simile cui non è stata applicata la misura preventiva in questione. Questo gruppo di controllo può non essere sottoposto a misure preventive, oppure essere sottoposto a quelle comunemente applicate.

Gli esperimenti possono richiedere il confronto fra più gruppi. Per esempio, si possono valutare diversi dosaggi di un farmaco, oppure vi possono essere due o più aspetti o elementi in un programma preventivo nel qual caso ogni gruppo sperimentale può ricevere un differente elemento o una diversa combinazione di elementi. Gli esperimenti possono anche prevedere programmi più complessi, in cui ogni gruppo riceve successivamente più trattamenti, possibilmente con intervalli di interruzione del trattamento.

Gruppi di controllo randomizzati

Il modo più tradizionale per definire il gruppo dei soggetti trattati e dei controlli è quello di identificare un gruppo di individui idonei e di suddividerlo in modo casuale in due o più sottogruppi. Se solo il caso determina chi partecipa a un gruppo piuttosto che a un altro è possibile applicare i test classici di significatività statistica per verificare con quale probabilità i risultati osservati possano essere dovuti al caso. Inoltre la randomizzazione contribuisce a rendere uniformi il gruppo sottoposto a trattamento e quello di controllo relativamente a fattori, sia noti che sconosciuti, che potrebbero influenzare i risultati dello studio.

Normalmente l'assegnazione casuale ai gruppi dovrebbe essere fatta dopo che ogni individuo è risultato idoneo e ha accettato di partecipare allo studio, per ridurre al minimo le perdite successive da uno o più gruppi. La randomizzazione effettuata prima di conoscere la disponibilità dei soggetti, o prerandomizzazione, può permettere di ridurre i rifiuti a partecipare allo studio, poiché non si può chiedere ai soggetti il consenso senza che essi conoscano il trattamento a cui verrebbero sottoposti. Sfortunatamente però, questa strategia può accentuare il problema di un "crossing-over" cioè del passaggio dei soggetti al trattamento non assegnato. Il cambiamento di gruppo rende più difficile cogliere le differenze fra i trattamenti, perché per quanto riguarda l'analisi dei dati è necessario che tutti gli individui rimangano nel gruppo assegnato inizialmente, indipendentemente dal trattamento ricevuto. Per questo motivo la prerandomizzazione è stata poco usata e necessita di ulteriori verifiche.

Nel caso in cui sia importante che il gruppo dei trattati e quello dei controlli siano omogenei per quanto riguarda determinate caratteristiche che potrebbero influenzare i risultati, l'intera popolazione oggetto di studio può essere suddivisa, o stratificata, in sottogruppi, e ogni sottogruppo può poi venire a sua volta diviso in modo casuale in soggetti trattati e controlli. Per esempio, la stratificazione in sottogruppi d'età può essere necessaria per garantire che il gruppo dei casi e quello dei controlli abbiano una distribuzione per età simile.

Se, dopo la randomizzazione, il ricercatore vuole essere sicuro che alcune caratteristiche critiche, per le quali non è stata eseguita la stratificazione, siano simili nel gruppo trattato e in quello controllo, deve esaminare la distribuzione di queste caratteristiche nei due gruppi. Se tali caratteristiche differiscono in maniera apprezzabile, ciò significa che la randomizzazione non è stata fortunata. La randomizzazione può, se possibile, essere ripetuta, oppure i risultati dell'esperimento debbono essere analizzati in modo da tenere conto delle differenze riguardanti queste caratteristiche critiche.

Gruppi di controllo non randomizzati

Non sempre è possibile in uno studio sperimentale disporre di gruppi di controllo randomizzati; le ragioni sono spesso di tipo economico. Può succedere che i fondi stanziati non permettano di seguire nel tempo i gruppi, sia di controlli che di casi, della grandezza richiesta; oppure la maggiore sicurezza che questa procedura ideale comporta può non essere tale da giustificare i costi che richiede; inoltre potrebbe non essere disponibile un numero sufficiente di individui per i due gruppi.

Anche se vi fossero individui e denaro sufficienti, la scelta casuale potrebbe rivelarsi impossibile o non realizzabile nella pratica; è questo il caso di una misura preventiva che può essere applicata solo all'intera popolazione, come l'aggiunta di elementi all'acqua potabile, oppure quando i controlli, venuti a conoscenza delle misure preventive attraverso gli individui trattati o una campagna pubblicitaria, le adottano nella stessa misura dei trattati.

Se non viene utilizzato un gruppo di controllo randomizzato, sono disponibili standard di confronto alternativi. Un gruppo di controllo può essere scelto fra persone simili al gruppo sperimentale per età, sesso, occupazione e

classe sociale. Oppure, quando un programma di prevenzione viene attuato su un'intera comunità, si può usare una comunità simile non trattata come controllo.

Un'altra possibilità è quella di usare il gruppo sperimentale come controllo di se stesso. Ovvero, si effettua un confronto prima-dopo, che prevede un periodo di osservazione del gruppo sperimentale prima che venga applicato il programma di prevenzione. L'andamento della malattia in questo periodo viene confrontato con quello osservato dopo l'attuazione del programma. Analogamente, anche se possono non essere coinvolti esattamente gli stessi individui, una precedente esperienza sulla comunità, o su un'altra popolazione definita, può essere confrontata con ciò che capita dopo l'introduzione di misure preventive. Gli individui osservati nel periodo precedente sono chiamati controlli storici.

Anche quando esiste un gruppo distinto per il confronto, è utile un periodo di osservazione. Se, durante questo periodo emergessero differenze sistematiche fra i gruppi, se ne può tenere conto nel confrontarli dopo l'applicazione della misura preventiva.

Quando non vengono utilizzati gruppi di controllo randomizzati è importante tenere conto di possibili vizi e differenze fra i gruppi. Utilizzare lo stesso gruppo come controllo di se stesso può sembrare un approccio interessante, perché sembra eliminare virtualmente le differenze tra gruppi. Tuttavia, le osservazioni nel gruppo dei controlli e in quello sperimentale vengono effettuate in momenti diversi. Si corre perciò il rischio che, con il passare del tempo, si verifichino cambiamenti nel gruppo in studio che potrebbero essere erroneamente attribuiti alla misura preventiva, in realtà inefficace, oppure che potrebbero aver mascherato eventuali benefici reali.

La collaborazione dei soggetti in studio

Molte misure preventive richiedono la cooperazione o l'attiva partecipazione degli individui che vengono studiati. Le valutazioni sperimentali di queste misure devono tenere conto del fatto che spesso molti individui non collaborano. Anche dopo aver acconsentito inizialmente a partecipare allo studio può capitare che una persona per varie ragioni voglia, a un certo punto, uscirne. Inoltre, nel gruppo dei trattati ci saranno quelli che non si attengono al trattamento o che vi si attengono solo in parte. Analogamente, nel gruppo dei controlli ci sarà qualcuno che, di nascosto, si procurerà, per esempio, il farmaco somministrato ai soggetti trattati.

L'analisi dei risultati non deve limitarsi agli individui che cooperano in ogni gruppo, perché essi rappresentano un sottogruppo selezionato, spesso caratterizzato da un grado di istruzione abbastanza elevato, da un elevato livello socio-economico, da una maggiore consapevolezza dei problemi sanitari e da abitudini di vita più salutari. Inoltre, nel momento in cui la misura preventiva venga adottata, essa deve essere applicata al mondo reale, dove ci sono moltissimi individui che non collaborano. Pertanto, il confronto più importante è quello tra l'intero gruppo sperimentale e l'intero gruppo di controllo, ossia la cosiddetta intention-to-treat analysis. Questo rappresenterà la migliore stima dei benefici che si possono ottenere dall'applicazione della misura preventiva.

Qualora si confrontino, in aggiunta, i soggetti che cooperano, l'analisi dovrebbe tener conto, nei limiti del possibile, delle loro caratteristiche specifiche e i risultati dovrebbero essere interpretati con cautela.

Esperimenti in cieco

I soggetti sperimentali dovrebbero essere tenuti all'oscuro, se possibile, della loro condizione di soggetto trattato o di controllo. Solo in questo modo è possibile evitare comportamenti che potrebbero impedire un'esatta valutazione degli esiti. Spesso però, la natura della terapia rende impossibile tenere i soggetti "ciechi" riguardo alla loro appartenenza al gruppo dei trattati o a quello dei controlli.

E' più importante che la valutazione degli esiti avvenga in condizioni di "cieco". In tutti i casi in cui ciò sia possibile, i medici o coloro che giudicano se le conseguenze di una malattia si siano manifestate o meno e in che grado, dovrebbero essere all'oscuro, cioè non sapere, se l'individuo in esame sia un soggetto trattato piuttosto che un controllo. L'uso di test e di criteri diagnostici obiettivi aiuta a prevenire vizi in favore del gruppo dei trattati o di quello dei controlli.

Talvolta, anche quando gli esperimenti sono progettati "in cieco", i soggetti in studio o gli sperimentatori vengono a conoscenza del loro stato. Se il trattamento prevede l'uso di farmaci, effetti collaterali caratteristici possono svelare il gruppo di appartenenza. Inoltre, all'insaputa del ricercatore, il personale medico che ha in cura i pazienti può venire a conoscenza del codice o di altre informazioni che permettono di identificare i trattati e i controlli.

Per questi motivi gli esperimenti "in cieco" sono spesso auspicabili ma meno frequentemente realizzati e, come in qualsiasi tipo di studio, è necessario valutare attentamente i metodi e i risultati per evitare possibili vizi.

Frequentemente ci si trova di fronte al termine doppio-cieco. Alcuni autori lo usano riferendosi a esperimenti dove sia l'assegnazione degli individui ai due gruppi sia la valutazione dei risultati sono fatti "in cieco". Altri lo impiegano nel riferirsi a esperimenti in cui né i pazienti né i medici sanno a quale dei gruppi appartengono i soggetti sperimentali.

Considerazioni sulla dimensione del campione e analisi sequenziale

Alcuni metodi statistici permettono di calcolare a priori quale deve essere l'ampiezza dei gruppi di controllo e dei trattati per ottenere risposte della precisione voluta. In generale quanto maggiore è il numero degli individui, tanto maggiore è la probabilità che i risultati dell'esperimento siano accurati e non soggetti a variazioni casuali.

Il fatto che sia consigliabile disporre di un elevato numero di individui è controbilanciato dalla considerazione pratica dei costi e delle difficoltà. Sorgono anche problemi etici nel decidere l'ampiezza del campione, infatti maggiore è il numero degli individui che partecipano allo studio, tanto più saranno le persone che ricevono un trattamento peggiore, sia che risulti più efficace il regime sperimentale che quello di controllo.

In alcuni casi gli individui entrano in uno studio sperimentale in periodi di tempo diversi; così, i risultati relativi a coloro che sono entrati nello studio per primi sono noti prima della conclusione dello studio. Si è tentati, allora, di valutare i risultati nei primi soggetti e di porre fine all'esperimento qualora sia evidente una differenza fra il gruppo dei controlli e quello dei trattati. Sfortunatamente, questi primi risultati non sono dotati dell'accuratezza originariamente prevista e stabilita per l'esperimento. Bloccare un esperimento a questo punto può sembrare economicamente ed eticamente giustificabile ma, a meno che le differenze osservate non siano inequivocabili e sostanziali, il ricercatore in seguito potrebbe pentirsi di aver basato le sue conclusioni su dati incompleti. Tuttavia, le differenze fra i trattati e i controlli possono essere talvolta molto maggiori di quelle attese in origine e dimostrabili in modo accurato. I ricercatori non vorranno certamente continuare l'esperimento una volta sicuri di trovarsi in questa situazione.

L'analisi sequenziale è un metodo statistico che permette di porre fine a un esperimento non appena sia stato ottenuto un risultato della precisione desiderata. Il risultato del confronto di ogni coppia di individui, un trattato e un controllo, viene valutato non appena è disponibile e aggiunto quindi ai risultati ottenuti precedentemente. Il criterio per decidere in favore del trattamento sperimentale o di quello di controllo è stabilito preventivamente, insieme al grado di accuratezza desiderato. Il confronto di un numero relativamente piccolo di coppie può mettere in luce differenze sostanziali che permettono di decidere in favore dell'uno o dell'altro.

Se ciò non si verifica, è necessario continuare ad aggiungere i risultati dell'analisi di nuove coppie fino a quando si raggiungono i criteri prefissati, o fino a che non diventa chiaro che non esistono differenze apprezzabili. Comunque, non appena è possibile trarre delle conclusioni, l'esperimento viene interrotto.

Consideriamo uno studio che valuti un farmaco per l'emicrania (Trial A). Supponiamo che alla fine del trial l'emicrania sia presente nel 30% del gruppo di controllo e solo nel 5% del gruppo trattato (rischio di 0.05).

Sebbene i concetti di Rischio, RR, RRR e RRA siano abbastanza chiari, essi hanno lo svantaggio di essere non dimensionabili e di non avere una diretta connessione con la situazione clinica, non chiarendo quindi immediatamente, a colpo d'occhio, l'efficacia del trattamento. Infatti il RR e RRR sono uguali nei 3 trials.

Quindi, oltre a questi parametri, vanno considerati altri parametri in modo tale da avere una visione più completa e dettagliata dei dati. Uno di questi è il NNT, che è semplicemente il reciproco del RRA.

	Trial A		Trial B		Trial C	
	Trattati	Controlli	Trattati	Controlli	Trattati	Controlli
Pazienti (n)	100	100				
Incidenza (i)	5	30				
RA (i/n)	0.05 (a)	0.30 (b)	0.14 (a)	0.8 (b)	0.00017 (a)	0.001 (b)
RR (a/b)	1.17 (c)		1.17 (c)		1.17 (c)	
RRR ((b-a)/b oppure 1-c)	0.83 (d)		0.83 (d)		0.83 (d)	
RRA b-a	0.25 (e)		0.66 (e)		0.00083 (e)	
NNT 1/e	4		1.5		1204	

RR: Rischio Relativo

RRR: Riduzione del Rischio Relativo

RRA: Riduzione del Rischio Assoluto

NNT: Numero Necessario da Trattare

PERCHE' RANDOMIZZARE

Per evitare il bias o vizio.

Negli studi statistici di associazione tra popolazioni diverse per verificare una correlazione statistica, si possono verificare diversi vizi o Bias che possono sovrastimare l'associazione o sottostimare.

Ci sono due tipi di Bias o vizi; vizi di selezione e vizi di informazione. A questi alcuni epidemiologi aggiungono un terzo o vizio di confondimento.

Il vizio di selezione si presenta quando gli individui studiati sono reclutati con modalità che possono aumentare o diminuire una associazione.

Per esempio, volendo trovare una un'associazione tra fumo e cancro del polmone conduco uno studio caso-controllo. Se prendo il campione di controllo da una popolazione di broncopatici, notoriamente fumatori, posso concludere che non vi è nessuna associazione tra fumo e cancro del polmone. Oppure se prendo come popolazione di controllo individui testimoni di Geova che notoriamente non bevono e non fumano, noterò una fortissima associazione. Quindi a seconda della selezione del campione di controllo posso trovare una sovrastima o una sottostima.

Un vizio di informazione può derivare quando viene usato un metodo improprio per la raccolta dei dati creando delle differenze fuorvianti nel tipo di informazioni raccolte da due o più gruppi diversi.

Per esempio, ci sarà un vizio di informazione se paragono due studi sulla obesità in cui uno ha raccolto i dati pesando i soggetti obesi e l'altro ha raccolto i dati chiedendo il peso da questionari.

Un vizio di confondimento si ha quando non si prendono in considerazione il sesso e l'età.

Per esempio se studio l'associazione tra un certo tipo di cancro e la forza in un campione posso trovare una associazione, ma se stratifico il campione in base al sesso non la trovo.

Errori comuni nell'assegnazione dei pazienti.

- assegnazione per gruppi di medici (diversa abilità clinica o chirurgica) o cliniche perché i pazienti ricevono trattamenti e assistenza diversi.
- assegnazione alternata a trattamenti diversi perché il medico può identificare il trattamento.
- assegnazione per numero di cartella clinica o numero del letto
- assegnazione per data di nascita perché essendo sistematica si presta a bias di selezione e inoltre esiste correlazione tra mese di nascita e morbilità

La randomizzazione semplice

La randomizzazione semplice (ossia l'equivalente di gettare una moneta), è la procedura di randomizzazione più elementare e probabilmente la più comunemente usata in pratica, ma non è sempre la scelta migliore..

Qualsiasi stratagemma per indovinare l'allocazione successiva è totalmente inutile nei confronti della randomizzazione semplice. La teoria statistica determina come conseguenza che il rapporto di assegnazioni a ciascun trattamento si avvicinerà all'obiettivo teorico man mano che il numero di pazienti cresce indefinitivamente.

Una randomizzazione non paritetica può venir consigliata in un gran numero di studi clinici in cui due trattamenti vengono comparati: nessuna seria perdita di efficienza risulta dall'assegnare una maggior proporzione di pazienti al trattamento di tipo sperimentale (a patto che il rapporto teorico tra trattamenti non sia superiore a 2:1), e vi sono due vantaggi - si ottiene una maggior esperienza clinica sul nuovo trattamento (in termini di tolleranza, tossicità, etc.), ed è possibile avere una stima più precisa dei suoi effetti.

Uso delle tavole dei numeri casuali.

Il primo passo nel determinare l'assegnazione randomizzata è stabilire una corrispondenza tra numeri e trattamenti. Ammettiamo di assegnare 20 persone al trattamento A e B e che i numeri dispari corrispondano al trattamento A ed i numeri pari al trattamento B.

Il secondo passo è determinare un modo sistematico e conveniente di leggere le tabelle dei numeri casuali (random) scegliendo un punto di partenza. Con una matita si sceglie un punto della tabella e vediamo a quale numero corrisponde; esempio il 50 e assegnando ad ogni persona un numero della colonna.

TAVOLA DEI NUMERI CASUALI

25	85	52	40	80	50	80	78	58	42	11	31	85	77	77	25	16	09	54	37
58	73	38	58	78	92	12	38	43	41	31	77	97	30	33	45	00	17	60	35
56	04	44	17	00	38	61	37	54	84	38	54	05	96	18	96	20	83	65	29
96	22	27	19	23	83	09	18	22	67	17	31	63	08	80	18	68	08	47	88
29	70	86	38	78	04	51	58	31	92	12	30	98	81	53	09	07	63	23	16
83	86	48	37	00	91	51	91	62	88	04	62	94	63	12	46	51	12	55	22
24	10	43	44	80	33	91	59	46	71	46	72	33	99	13	16	51	34	84	49
26	45	22	19	59	42	70	02	*50	23	78	14	24	19	86	33	37	11	65	36
59	58	24	97	89	51	48	61	85	92	96	83	36	62	06	90	17	23	92	17
14	19	83	76	52	64	91	14	70	66	74	05	47	54	26	81	32	14	91	27
27	70	78	05	51	37	28	77	39	03	63	18	66	74	90	45	23	47	49	80
98	13	78	10	97	38	27	61	27	91	81	70	55	28	35	37	45	76	48	82
94	08	72	66	37	34	44	45	32	26	04	37	80	55	29	88	65	99	63	64
49	75	05	70	91	47	21	85	45	96	57	06	08	49	03	91	19	89	87	10
17	55	25	70	35	57	96	81	90	91	91	05	56	31	55	22	15	63	67	57
57	83	92	42	51	67	49	06	19	44	23	43	39	33	65	52	69	76	98	66
54	36	96	45	91	71	43	43	93	04	71	90	40	81	82	26	56	84	31	32
95	36	83	47	04	07	56	30	59	82	14	82	28	69	27	22	19	47	18	08
61	12	39	28	72	79	97	30	82	43	58	53	40	49	99	71	09	76	60	35
26	27	83	09	88	15	45	34	19	86	67	13	18	50	45	33	02	48	34	58
45	24	50	85	51	89	13	09	85	23	33	26	92	73	47	94	81	86	25	94
44	77	24	50	96	30	18	51	79	05	26	81	87	43	50	71	99	55	32	09
19	79	28	20	23	65	41	96	95	76	81	06	28	11	36	66	67	15	33	42
55	17	44	90	96	22	30	63	61	12	03	95	61	40	29	69	40	05	50	67
73	57	60	29	70	47	76	17	78	45	19	42	58	24	02	39	16	40	86	60
89	02	98	00	17	62	52	94	92	56	80	00	00	26	31	95	24	34	84	41
82	02	00	81	89	86	31	61	31	23	38	70	16	15	93	06	97	44	33	63
52	01	93	94	50	55	19	39	99	45	33	41	40	40	71	54	95	11	19	13
64	80	82	89	45	39	32	75	89	53	80	65	54	77	23	81	89	46	93	31
78	83	74	73	25	33	78	96	66	49	40	11	35	35	17	50	41	41	48	02
16	71	31	07	82	88	89	96	44	21	38	20	22	77	23	29	83	14	38	06
98	90	44	87	42	03	23	23	87	14	30	08	25	17	46	51	03	58	46	62
34	95	90	75	63	46	57	53	13	59	91	93	58	51	43	59	78	88	37	65
30	27	22	59	07	85	95	97	61	29	31	56	45	21	25	71	99	16	31	37
00	08	48	54	49	72	89	48	59	49	15	41	70	48	69	97	49	77	29	22
33	57	45	38	37	68	28	69	03	29	50	80	59	48	51	60	95	08	65	40
54	34	43	44	14	00	94	50	57	57	68	15	08	20	25	55	19	37	75	39
74	21	68	80	20	75	39	51	58	77	39	47	93	74	54	29	25	18	18	59
65	09	82	45	75	80	58	91	62	39	49	97	01	59	90	83	42	44	47	34
10	43	48	61	60	65	96	45	17	33	03	50	79	26	85	72	61	01	67	75
01	94	95	31	34	40	15	77	57	25	94	76	77	95	34	61	57	80	84	91
57	72	84	03	52	33	45	45	54	58	76	24	11	70	47	64	31	34	00	52
85	89	47	95	04	23	30	42	91	63	50	27	31	88	37	53	93	46	14	34
21	43	25	72	26	18	52	81	68	53	65	65	62	77	29	63	81	17	21	73
73	62	52	21	84	78	63	17	45	28	35	53	60	79	80	43	87	65	80	44
55	91	18	37	95	71	45	13	16	10	65	76	74	52	87	59	10	93	00	48
26	40	72	27	23	11	80	68	17	82	54	90	09	34	16	03	92	54	62	71
28	53	58	92	62	13	56	50	36	18	96	48	96	31	03	67	94	99	66	37
39	70	30	58	87	42	52	62	06	97	20	58	52	63	22	29	13	57	34	10
75	78	85	97	01	92	56	74	85	83	02	12	07	93	39	86	31	22	41	32

Tutte le fasi andrebbero documentate, cosicché la randomizzazione possa essere seguita o continuata da un altro collega. La randomizzazione fatta gettando una moneta o tirando un dado non è raccomandabile perché non può venir controllata o riprodotta.

Il terzo passo è eseguire le assegnazioni dei trattamenti secondo il sistema definito sopra. Cosicché, i pazienti dall'1 al 20 verranno assegnati nel modo seguente:

Paziente	Numero casuale	Assegnazione
1	50	trattamento A
2	85	trattamento B
3	70	trattamento A
4	39	trattamento B
5	27	trattamento B
6	32	trattamento A
7	45	trattamento B
8	90	trattamento A

9	19	trattamento B
10	93	trattamento B
11	59	trattamento B
12	82	trattamento A
13	19	trattamento B
14	85	trattamento B
15	79	trattamento B
16	95	trattamento B
17	61	trattamento B
18	78	trattamento A
19	92	trattamento A
20	31	trattamento A

L'assegnazione, sopra riportata, porta alla conseguenza che solo otto pazienti su 30 ricevono il trattamento A . Quindi questo tipo di assegnazione verrebbe rifiutata e ripetuta fino a che si otterrebbe una assegnazione più bilanciata, oppure verrebbe usato un metodo di randomizzazione ristretta. Per ripetere la randomizzazione semplice occorre continuare a leggere la colonna fino in fondo e poi dall'inizio della successiva.

Quando non è vantaggioso usare la randomizzazione semplice?

- Quando la dimensione finale dello studio è inferiore ai 100 pazienti.
Si formano studi sbilanciati come sopra. Sebbene la randomizzazione semplice possa andar bene per lo studio nel suo complesso, il numero dei trattamenti o la distribuzione di determinate caratteristiche dei pazienti possono risultare decisamente sbilanciati alla prima e, talvolta, anche a successive analisi preliminari, il che può portare, di conseguenza, a qualche difficoltà. Il problema dello sbilanciamento di qualche variabile relativa ai pazienti nel corso di analisi preliminari può verificarsi anche con altri schemi di randomizzazione.
- Negli studi multicentrici.
- Se si prevedono analisi intermedie.

Gli svantaggi insiti nella randomizzazione semplice dimostrano la necessità di disporre di metodi di randomizzazione ristretta e stratificata.

La randomizzazione ristretta

La randomizzazione ristretta è consigliabile quando ci si voglia assicurare che il numero di pazienti assegnati a ciascun trattamento sia approssimativamente equidistribuito nello studio nel suo complesso in importanti sottogruppi di pazienti, od in entrambi. I metodi relativi, in ordine di complessità crescente, sono blocchi randomizzati permutati, randomizzazione stratificata e minimizzazione.

Il metodo dei blocchi randomizzati permutati è facile da usare. Nella globalità dello studio, esso garantisce che il numero dei pazienti assegnati a ciascun trattamento sia uguale dopo che ogni blocco o altro gruppo di pazienti è entrato nello studio. La stratificazione per un fattore o, in certi casi, due, che notoriamente influenzano la prognosi costituisce una salvaguardia contro uno sbilanciamento casuale tra i diversi gruppi di trattamento nei confronti di una variabile importante; per esempio, le dimensioni delle caverne tubercolari o il coinvolgimento dei linfonodi ascellari nel caso del tumore della mammella. La stratificazione, in particolar modo in studi clinici di dimensioni limitate è raccomandabile nella misura in cui la randomizzazione resta semplice da effettuare. Un metodo per effettuarla , costituito semplicemente dal preparare liste separate per ciascuno strato col sistema dei blocchi randomizzati permutati, ossia, consultare liste differenti a seconda del grado di estensione delle caverne di pazienti con tubercolosi.

La randomizzazione stratificata può essere utile in studi multicentrici ove sia importante evitare sbilanciamenti nell'assegnazione dei trattamenti nei singoli ospedali oltre che nello studio nel suo complesso. Un sistema per ottenere ciò nel caso in cui vi siano numerosi centri consiste nel preparare una lista di randomizzazione per lo studio nel suo complesso, monitorare lo sbilanciamento nei singoli ospedali, ed intervenire per riequilibrare l'assegnazione all'interno dei singoli ospedali prima che sia troppo disomogenea.

La minimizzazione, come dice la parola stessa, è un metodo di assegnazione randomizzata che minimizza lo sbilanciamento marginale nel numero di pazienti assegnati a diversi trattamenti in riferimento a parecchi (due o più) fattori che notoriamente influenzano la prognosi, uno dei quali può essere l'ospedale di trattamento nel caso di uno studio multicentrico. Questo metodo evita le limitazioni della randomizzazione stratificata, ma ha uno scopo del tutto simile. Funziona in questo modo: una misura di sbilanciamento viene calcolata su tutta una serie di fattori prognostici che descrivono il nuovo paziente, che è poi con maggiore probabilità, ma non invariabilmente, assegnato al trattamento che minimizza lo sbilanciamento globale.

I disegni randomizzati con consenso del paziente hanno diversi scopi: (a) limitare il numero dei pazienti ai quali debba venir data una piena e spesso difficoltosa spiegazione dello scopo di uno studio clinico randomizzato; e (b) incoraggiare la partecipazione dei clinici agli studi. Alcuni medici, infatti, temono che il consenso informato mini la fiducia e la speranza dei propri pazienti. Qualcuno ha proposto pertanto che tutti i pazienti teoricamente eleggibili per lo studio venissero randomizzati ad un gruppo in cui si cercava il consenso, e ad un altro in cui non lo si cercava. Questo secondo gruppo riceve semplicemente il trattamento standard, mentre al primo gruppo viene chiesto di dare un consenso informato al trattamento sperimentale. Il confronto viene poi fatto tra i gruppi come sono stati randomizzati, sebbene anche il gruppo in cui si è cercato il consenso abbia una certa proporzione di pazienti che hanno ricevuto il trattamento standard - poiché così hanno deciso dopo che lo studio è stato loro spiegato nei dettagli, o poiché il medico ha scelto di non metterli neppure di fronte ad una spiegazione traumatica.

Blocchi randomizzati permutati

Il metodo dei blocchi randomizzati permutati funziona bene a patto che il medico non riesca ad intuire la lunghezza dei blocchi, e non possa identificare i trattamenti che sono stati assegnati ai pazienti precedenti nel blocco. Se invece può identificare i primi tre trattamenti assegnati e sa che la lunghezza del blocco è quattro, allora capisce automaticamente che l'ultimo paziente nel blocco deve ricevere il trattamento che equilibra l'assegnazione. Il bias di selezione diventa allora un problema, specialmente se la lunghezza del blocco è limitata (ugual numero di pazienti per trattamento dopo ogni due, quattro o sei pazienti entrati).

I trattamenti A e B vengono assegnati in modo tale che per ogni blocco di 4 pazienti sia un ugual numero di pazienti per ogni trattamento. L'assegnazione per ciascun blocco di 4 pazienti deve pertanto essere una delle sei combinazioni seguenti:

1	A	A	B	B
2	A	B	A	B
3	A	B	B	A
4	B	B	A	A
5	B	A	B	A
6	B	A	A	B

Randomizzazione a blocchi permutati: lunghezza del blocco 4 pazienti. (a) vi sono in tutto 24 sequenze diverse di 3A e 3B corrispondenti ad una lunghezza del blocco di 6. (b) le tavole delle permutazioni randomizzate dovrebbero venir usate se vi sono troppe sequenze diverse da elencare, esempio quando le dimensioni di ciascun blocco superano i sei pazienti. (c) in studi unicentrici occorre evitare blocchi di dimensioni limitate, esempio occorrerà scegliere una lunghezza per blocco di 10 o più pazienti, o variare in modo casuale la lunghezza dei blocchi.

Nell'esempio, il bias di selezione potrebbe influenzare la decisione del trattamento in un terzo dei pazienti entrati in studio. Chiaramente, lo statistico non dovrebbe dire al clinico qual'è la lunghezza del blocco per rendere più difficoltosa la scoperta delle regole.

Stratificazione

Perché una stratificazione eccessiva è dannosa?

- Scoraggia la partecipazione allo studio
- Porta come conseguenza ad un numero eccessivo di strati con troppo pochi pazienti

- La complessità amministrativa è essa stessa causa di errori.

Vale la pena perciò di considerare soltanto la stratificazione per variabili che si sappia o si sospetti influenzino la prognosi, che, di solito, sono poche. Una stratificazione eccessivamente zelante è un problema - i livelli dei diversi fattori devono essere moltiplicati (non addizionati) per dare il numero totale dei sottogruppi. Persino tre soli fattori prognostici - quali le dimensioni del tumore, il coinvolgimento dei linfonodi ascellari e lo stato di menopausa nel carcinoma della mammella, ciascuno su tre livelli - dimensioni nel tumore: < 2 cm, 3-4 cm, >= 5 cm; coinvolgimento dei linfonodi ascellari: non coinvolti, mobili, o fissi, stato di menopausa: premenopausa, in menopausa o in post-menopausa - comportano $3 \times 3 \times 3 = 27$ sottogruppi di pazienti. Per ciascuno di essi, occorre consultare una lista separata di randomizzazione ristretta. Ed un problema ancora maggiore è costituito dal fatto che è improbabile la distribuzione dei pazienti sia uniforme, cosicché molti strati includeranno così pochi pazienti che la procedura di randomizzazione ristretta non potrà neppure essere in realtà attuata - per fare un esempio, bilanciare il numero dei trattamenti ogni sei pazienti entrati non è possibile nel caso di strati con meno di sei pazienti. E di conseguenza il numero dei trattamenti per strato non potrà essere omogeneo, anche se lo studio nel suo complesso necessita che lo sia. Una stratificazione eccessiva è pertanto controproducente. Inoltre, un aggiustamento può venir attuato poi retrospettivamente nel corso dell'analisi dei dati per tener conto di moderate differenze tra gruppi di trattamento in relazione ad una variabile - l'età al menarca, per esempio - che non era stata in precedenza considerata rilevante come fattore prognostico.

Minimizzazione

Lo schema che segue mostra l'assegnazione di trattamenti fino a questo punto: 60 pazienti con carcinoma della mammella sono state randomizzate a mastectomia semplice + radioterapia od a mastectomia radicale. La paziente numero 61 è in premenopausa, ha un tumore, delle dimensioni di 5 cm e linfonodi ascellari positivi. Quale assegnazione di trattamento conduce al minor sbilanciamento per ciò che riguarda i fattori prognostici rilevanti.

- Mastectomia semplice + radioterapia

- La paziente 61 è assegnata a questo trattamento con probabilità maggiore di 1/2 ma minore di 1 - il che significa che nello stabilire l'assegnazione viene favorita la mastectomia semplice + radioterapia.

Bilanciamento dei trattamenti per diversi fattori prognostici

Trattamento	pre-menopausa	post-menopausa	>=2	3-4	>=5	negativi	positivi
Mastectomia semplice + radioterapia (n=30)	8	22	4	14	12	17	13
Mastectomia radicale (n=30)	7	23	5	12	13	14	16
Numero totale delle pazienti	15	45	9	26	25	31	29

La paziente 61 è in pre-menopausa, ha un tumore di 5 cm e linfonodi positivi

Il secondo schema mostra che di 15 pazienti in pre-menopausa trattate fino a questo momento, sette hanno subito una mastectomia radicale. Assegnando anche la paziente 61 al gruppo con mastectomia radicale, le proporzioni si sarebbero perfettamente equiparate nei due gruppi di trattamento. D'altra parte, l'assegnazione a mastectomia semplice + radioterapia viene preferita al fine di minimizzare lo sbilanciamento tra i diversi gruppi di trattamento in riferimento sia alle pazienti con una grossa massa primaria che con linfonodi positivi.

	<u>Assegnazione</u>		Assegnazione favorita per la paziente 61
	Mastectomia Semplice + Radioterapia	Mastectomia Radicale	
Pre-menopausa	8	7	Mastectomia radicale
Dimensione del Tumore ≥ 5	12	13	Mastectomia semplice + radioterapia
Linfonodi positivi	13	16	Mastectomia semplice + radioterapia

Una misura grossolana dello sbilanciamento complessivo potrebbe essere data dal numero di voti per e contro ciascun trattamento. La mastectomia semplice + radioterapia vincerebbe perché ha due voti. Ma questo tipo di misura può venir criticata perché non tiene conto del fatto che uno sbilanciamento di 13 verso 16 necessita maggiormente di venir corretto rispetto ad uno di 8 verso 7. Così se la paziente 61 venisse assegnata alla mastectomia radicale lo sbilanciamento complessivo sarebbe allora $(8-8) + (12-14) + (13-17) = -6$, rispetto a 0 se fossero scelte la mastectomia semplice + radioterapia. Anche questo criterio può favorire la mastectomia semplice+radioterapia. Ciò che succede di conseguenza è che la paziente 61 ha una elevata probabilità - per esempio 3/4 - di venir assegnata al gruppo con mastectomia semplice + radioterapia, ma potrebbe ciononostante essere ugualmente randomizzata alla mastectomia radicale - con probabilità 1/4 - il che accrescerebbe lo sbilanciamento. E' importante comunque mantenere un elemento di casualità - ossia una probabilità di assegnazione minore di 1 - per evitare il "bias" di selezione.

Disegno randomizzato con consenso

- Difficoltà nel rendere lo studio "doppio cieco"
- Soltanto i pazienti del gruppo cui viene chiesto il consenso sapranno di far parte di uno studio clinico controllato.
- I medici possono essere più persuasivi nel presentare le informazioni sul nuovo trattamento ad alcuni tipi di pazienti che non ad altri.

I disegni con consenso randomizzato hanno diversi limiti. Il primo è che è difficile fare in modo che siano in doppio cieco, poiché l'appartenenza ad un gruppo viene rivelata dal fatto che ad un paziente viene chiesto o no il consenso informato per il trattamento sperimentale. La seconda differenza tra i gruppi randomizzati è che i pazienti assegnati ad un trattamento sanno che il risultato del loro trattamento è di particolare interesse per il medico. Questo tipo di conoscenza può influenzare il grado di accettazione del trattamento o il modo in cui il paziente riferisce il suo stato, e così viziare il paragone tra i trattamenti. Un terzo problema si verifica quando i risultati vengono analizzati se la proporzione di pazienti che accettano il trattamento sperimentale è diversa nei vari sottogruppi. Una diversa proporzione di accettazione tra i pazienti non necessariamente deve riflettere un diverso grado di accettabilità del trattamento sperimentale, ma può dipendere dal grado di persuasività con cui i medici avevano presentato l'informazione sul nuovo trattamento. Questo problema diviene particolarmente complesso nel caso di terapie che sono di fatto superiori a quelle tradizionali in alcuni sottogruppi, ma inferiori in altri. Questa interazione potrebbe venir oscurata nei risultati dello studio se i clinici, con una corretta intuizione, assegnassero al trattamento sperimentale soltanto quei pazienti in cui si aspettano il maggior beneficio. In sede di analisi dei dati, il gruppo in cui si era cercato il consenso sarà quello con i risultati migliori, ma si noterà anche che in qualche sottogruppo una elevata proporzione di pazienti avrà rifiutato il trattamento sperimentale. Vi potrà non essere una chiara interpretazione delle ragioni del rifiuto; il rifiuto non è necessariamente un fattore a sfavore del trattamento sperimentale, ma lo può essere. I disegni randomizzati con consenso sono utili soltanto quando una proporzione consistentemente elevata di pazienti nel gruppo in cui si cerca il consenso accetti il trattamento sperimentale.

La randomizzazione telefonica

Un ufficio di randomizzazione centrale toglie ai singoli centri o medici l'onere di effettuare la randomizzazione e salvaguarda lo schema dai curiosi e dai furbi. La randomizzazione telefonica funziona bene a patto che (a) l'ufficio di randomizzazione sia coperto durante le ore di servizio prestabilito, in modo che non si facciano aspettare i medici quando telefonano per chiedere dettagli su un paziente includibile nello studio; e (b) che l'incaricato annoti il nome del paziente, l'ospedale, il nome del medico che effettua la chiamata e controlli l'eleggibilità del paziente per lo studio. L'assegnazione del gruppo di trattamento sarà poi comunicata e registrata. La data della telefonata corrisponde alla data della registrazione nello studio. I dettagli sono poi confermati per iscritto al medico, che da parte sua invierà una scheda più dettagliata per ogni paziente.

La variabilità intra-osservatori

Una scarsa ripetibilità implica una scarsa validità - come in epidemiologia, così anche nel caso degli studi clinici controllati. Qualora non sia possibile dare una risposta soddisfacente alla domanda: Quanto sono riproducibili i risultati ottenuti dallo stesso osservatore - per esempio con una seconda misurazione o su un altro campione - sono da preferirsi altri tipi di valutazione che misurino la stessa variabile ma che siano ripetibili. Questa regola si applica in laboratorio (selezione di diversi esami o tecniche di analisi), alla radiologia, agli ultrasuoni e ad altre tecniche strumentali, nonché alla registrazione clinica di segni o sintomi. Infatti, essi non danno informazione alcuna se vengono registrati in modo diverso dai medici in successivi esami del paziente.

Anche quando i problemi in questo settore vengono ristretti alle misurazioni ripetibili che siano anche valide - ossia, che misurino realmente ciò che intendono misurare - rimane il problema della variabilità tra i diversi osservatori. Un medico che rileva dei segni clinici in una patologia respiratoria, un neurologo che fa diagnosi di sospetta sclerosi multipla, un geriatra che valuta la riabilitazione dopo un episodio di ictus, ed un anestesista che valuta l'operabilità di un paziente, esprimono tutti un giudizio che potrebbe venir espresso diversamente da un altro medico. La variabilità tra diversi osservatori in uno studio clinico controllato può seriamente comprometterne i risultati. L'esempio più tipico in questo senso è dato dal caso in cui tutti i pazienti che ricevono il farmaco A sono valutati da un medico, mentre un altro medico valuta i pazienti che ricevono il farmaco B. Il risultato di uno studio di questo tipo è totalmente non informativo, poiché può venir in ogni caso attribuito a differenze tra gli osservatori piuttosto che tra i trattamenti.

Per evitare questo tipo di problema si potrebbe fare in modo che lo stesso osservatore valuti tutti i pazienti, ma una soluzione di questo tipo non sempre è possibile a causa di considerazioni legate all'efficienza nella conduzione di un reparto od agli impegni dei clinici stessi. Occorrerà assicurarsi almeno che lo stesso osservatore effettui le valutazioni successive su di un dato paziente, e quando vi sia più di un osservatore possa identificare quale di essi abbia valutato i diversi pazienti (usando un apposito codice per ogni osservatore).

Come può venir ridotta la variabilità tra i diversi osservatori?

- Usando un gruppo di esperti, od un singolo laboratorio di riferimento
- Confrontando diversi schemi di valutazione in uno studio pilota
- Cercando di scoprire in che modo l'osservatore possa fornire valutazioni sistematicamente erranee
- Fornendo un training formale al personale incaricato della valutazione
- Standardizzando le tecniche ed i criteri di valutazione.
- Cercando di stimare quanto sia rilevante la discrepanza tra osservatori
- Randomizzando i pazienti ai diversi osservatori
- Includendo un revisore indipendente nel gruppo di studio

Il problema della variabilità tra i diversi osservatori può venir affrontato nominando un gruppo di specialisti che possano verificare tutte le diagnosi: patologiche, radiologiche, scintigrafiche, etc. Idealmente, questa revisione dovrebbe venir fatta prima di randomizzare i pazienti ai diversi trattamenti, per evitare esclusioni a posteriori. I risultati patologici, ematologici o biochimici di uno studio, possono inoltre venir sottoposti al giudizio di un laboratorio di riferimento.

Tutti questi risultati dovrebbero ovviamente venir ricontrollati "in cieco", senza alcuna conoscenza dei trattamenti ricevuti.

Un altro tipo di approccio è confrontare i diversi schemi di valutazione in uno studio pilota per vedere quale schema offra il maggior consenso tra i diversi osservatori. Nel disegnare studi pilota di questo tipo, occorrerà ovviamente tener presente che quei pazienti che hanno mostrato un netto miglioramento od un apprezzabile peggioramento saranno identificati correttamente dalla maggior parte degli osservatori ed in base a ogni possibile criterio. Sono invece i pazienti che hanno mostrato una risposta parziale i più difficili da classificare. Un buon sistema di valutazione dovrà quindi portare ad un consenso anche su di essi.

Inoltre, è necessario ricordare che il consenso di per sé non è sufficiente: deve esserci un'evidenza che ciò che si valuta sia effettivamente un test diagnostico affidabile ad una misura di miglioramento clinico.

Il comprendere in che modo l'osservatore può commettere errori sistematici può ridurre la variabilità intra-osservatori. E' stato suggerito che un "bias" nei confronti di un abbassamento od aumento della pressione arteriosa di un paziente venga instaurato dal creare una divisione arbitraria tra pressione normale od elevata - e lo stesso fenomeno può essere causato dalla conoscenza di dati fattori di rischio, quali l'obesità, o dallo stabilire una soglia per l'inclusione in uno studio clinico controllato. Un altro fattore che può introdurre errori di misurazione è che talvolta il rilevatore può preferire valori pari quando degli strumenti, quali manometri o termometri, hanno scale che riportano soltanto i valori pari, o cifre che terminano con 0 e 5.

La variabilità intra-osservatori può essere limitata da procedure di standardizzazione - ossia aderendo a particolari protocolli, comprendenti ad esempio regole per valutazioni cliniche e selezione di test diagnostici, specificazione del tipo di reagenti da usare in esami di laboratorio, e liste di domande formulate con cura ed in un determinato ordine. La standardizzazione delle misurazioni della pressione arteriosa è stata ben illustrata, ma non adeguatamente riportata in pubblicazioni mediche. Occorre riportare il braccio usato, il tipo di apparecchio, il punto di registrazione della pressione diastolica, il numero di letture per misurazione, ecc.

Vi sono altre soluzioni possibili al problema della variabilità intra-osservatori. Assegnando in modo randomizzato i pazienti ai diversi osservatori, tutti i differenti tipi di pazienti dovrebbero risultare proporzionalmente rappresentati nel campione che viene valutato da ciascun osservatore. Ogni scheda dovrebbe avere un codice per il rilevatore (o la sua firma), per poter identificare quale osservatore ha valutato ciascun paziente; inoltre, le valutazioni seguenti su di un particolare paziente dovrebbero venir effettuate dallo stesso osservatore che ha registrato i primi dati del paziente, onde evitare un possibile confondimento tra effetti degli osservatori e dei trattamenti.

Gli studi clinici in doppio cieco

Che cosa si intende per studio clinico controllato in cieco semplice e in doppio cieco.

Uno studio clinico controllato è in doppio cieco quando né il medico né, il paziente possono identificare il trattamento assegnato.

Se l'identificazione del trattamento è nascosta soltanto a colui che deve valutarlo, lo studio viene definito in "cieco semplice".

Fare in modo che né il medico né il paziente possano identificare il trattamento assegnato è il modo ideale per evitare valutazioni caratterizzate da insidiosi errori sistematici (da parte dell'osservatore), o diverse descrizioni (da parte del paziente) dei risultati dei trattamenti.

Talvolta, la risposta al trattamento è inequivocabile (morte o sopravvivenza), cosicché l'uso del "doppio cieco" è una finezza del disegno piuttosto che un suo elemento essenziale. Più spesso i trattamenti - ad esempio la trazione per i dolori dorsali - sono tali che escogitare un "doppio cieco" è praticamente impossibile, sebbene vi siano esempi ingegnosi ed audaci di mascheramento dei trattamenti. E consigliabile, ovviamente, che i ricercatori descrivano con cura come è stato ottenuto il "cieco", o che spieghino chiaramente perché uno studio in cieco non sia fattibile.

In quali circostanze è importante che lo studio sia un semplice o doppio cieco?

In tutti i casi in cui l'essere a conoscenza del trattamento somministrato al paziente potrebbe introdurre errori sistematici nella valutazione della risposta da parte del medico, il trattamento dovrebbe essere almeno in cieco semplice. Inoltre, se la risposta del paziente al trattamento o le sue modalità di riferire la risposta possono venir influenzate dal conoscere quale trattamento riceve, lo studio dovrebbe essere in doppio cieco.

Errori nel registrare la data di morte sono inusuali. Errori sistematici nel riportare il tempo di insorgenza di metastasi sono tuttavia possibili, se un gruppo di trattamento è, magari involontariamente, seguito più regolarmente degli altri, o se i metodi usati per evidenziare le metastasi variano tra i diversi gruppi di trattamento.

Pregiudizi (in favore o contro) un particolare trattamento possono portare a riferire in modo inconsciamente errato la storia clinica, la pressione sanguigna, gli esami di laboratorio o radiografici, e molti altri tipi di misurazione.

Quali segni-spia possono rivelare il tipo di trattamento?

Gli indicatori-spia che possono rivelare i diversi trattamenti comprendono il loro sapore, odore o forma generale. Altri fattori di rilievo sono la forma ed il colore dei farmaci stessi o delle loro confezioni. La frequenza di somministrazione - ad esempio due oppure tre volte al giorno - o la tossicità sono talvolta sufficienti per distinguere i trattamenti. Inoltre, in uno studio in cieco semplice il paziente può semplicemente dire al medico al radiologo che tipo di trattamento ha ricevuto.

Ovviamente, un confronto di trattamenti in "cieco" non è sempre possibile..

L'organizzazione di uno studio

Dosi fisse o variabili

Si dovrebbe resistere alla tentazione di usare dosi fisse quando è necessaria una variazione considerevole dei dosaggi nei pazienti per ottenere un effetto farmacologico. Per esempio, nei diabetici la dose viene adattata alle necessità del paziente, e così, si dovrebbe fare anche in uno studio clinico controllato, altrimenti si rischia di dare risposte altamente attendibili a problemi insignificanti. Negli studi clinici controllati che realmente abbiano dei fini pratici i confronti dovrebbero essere fatti tra trattamenti che siano applicati in modo ottimale. Solo così si può affrontare un problema clinico rilevante. In tale situazione, il confronto a doppio cieco dovrebbe essere abbandonato e la rilevanza clinica avere chiaramente la priorità. Bisogna stabilire una valutazione indipendente "in cieco" quando la differenza tra i trattamenti non può realisticamente essere nascosta ai pazienti.

Quando in uno studio clinico controllato è giustificata la inclusione di un campione con placebo?

Uno studio clinico controllato prevede che il campione di controllo debba essere un trattamento comunemente accettato per la patologia in studio. Solo allora il risultato dello studio risponde ad un problema clinico rilevante. Il nuovo trattamento è migliore di un metodo standard? La dimostrazione che un nuovo farmaco è migliore di nessun trattamento può interessare le autorità preposte al controllo sui farmaci, ma non ha nessuna implicazione pratica per la cura dei pazienti.

L'includere un gruppo non trattato, un gruppo con un trattamento inattivo, od entrambi, negli studi che propongono un quesito terapeutico pratico - cioè, di tipo prettamente pragmatico - può essere giustificato solo se non vi è nessun trattamento riconosciuto e sperimentato per quella patologia. Se inoltre viene sospettato un effetto psicologico aspecifico o psicofisiologico od è probabile un errore sistematico di osservazione, dovrà essere previsto un gruppo trattato con placebo insieme od in alternativa ai pazienti non trattati.

Il non dare coscientemente un trattamento efficace è inammissibile a meno che i pazienti non diano un consenso pienamente informato; inoltre, ciò non deve andare contro l'etica professionale (condizioni molto rare). Parimenti inammissibile è l'accusa che la sola giustificazione per molti degli studi con placebo che vengono pubblicati sia la rapidità, con la quale si possono soddisfare gli statistici.

Ordine della valutazione

Lo stesso clinico prima intervista il paziente e poi compie l'endoscopia. Le informazioni sui sintomi e le impressioni dei pazienti su quanto sia stata efficace la terapia, possono così condizionare il risultato dell'endoscopia. Il clinico può cioè osservare ciò che si aspetta di osservare. Un metodo migliore è che un medico raccolga la storia clinica del paziente ed un altro, che non ha letto la cartella clinica esegua l'endoscopia, così che il secondo accertamento sia indipendente dal primo.

Due altri errori sono il cambiare colui che fa la valutazione, ed il tempo di valutazione. Il primo, e normalmente anche il secondo aspetto, dovrebbero essere già discussi nel protocollo dello studio. In uno studio sulla ipertensione essenziale, i pazienti furono esaminati ad un'ora prefissata ed istruiti a prendere le loro compresse nello stesso momento ogni giorno. In questo stesso studio, la pressione doveva sempre venir misurata dallo stesso osservatore e con una procedura definita precisamente.

L'aderenza (compliance) ai protocolli

Il semplificare il regime terapeutico, a condizione che questo non vada a scapito dell'attività farmacologica, riduce al minimo gli errori di comprensione e di memoria. Può essere di aiuto prescrivere il farmaco una sola volta al giorno, o sincronizzarne l'assunzione con l'ora del pasto od altre attività regolari. L'istruire i pazienti prima di dimetterli, a come usare i farmaci - per esempio, gli inalatori -, il dare ricette ed istruzioni facili da leggere e confezioni adatte - per esempio, confezioni-calendario - o schede-diario, rendono più semplice per il paziente rispettare lo schema di trattamento prescritto. Un gusto od un odore sgradevole possono diminuire l'accettabilità della terapia.

Il grado di aderenza al trattamento può essere monitorato chiedendo ai pazienti di portare i farmaci con loro per controllo quando tornano all'ospedale e chiedendo loro di restituire tutti quelli non usati. Naturalmente i pazienti possono facilmente gettare i farmaci non usati, se vogliono. Ma il risultato anche di un controllo così semplice può essere sorprendentemente. Un controllo più attendibile è l'analisi di campioni di sangue e di urine. In uno studio sull'uso di aspirina versus placebo nella prevenzione della mortalità dopo infarto del miocardio furono compiute visite domiciliari aggiuntive e non preventivate e vennero presi campioni di urine per misurare il salicilato.

I risultati suggerirono che il tasso di aderenza ai trattamenti era almeno del 72% (adeguata considerazione fu data ai tassi dei falsi positivi e falsi negativi).

Cosa si può fare per ridurre al minimo la perdita al follow-up dei pazienti in uno studio clinico controllato?

Ad un esame retrospettivo, la proporzione di studi nei quali i pazienti vengono persi al follow-up o si ritirano volontariamente è alta. Tali esclusioni di pazienti da uno studio clinico controllato possono essere ridotte al minimo (a) da una precisa valutazione prima della registrazione della eleggibilità del paziente ad entrare nello studio; (b) dalla restrizione della popolazione in studio (e di conseguenza anche delle implicazioni dello studio), ai pazienti per cui le visite di controllo non siano una seria imposizione sia per problemi familiari che fisici; (c) da un tempestivo invio di promemoria quando vengono saltati appuntamenti.

E inoltre responsabilità del coordinatore dello studio conoscere il prima possibile gli errori nella procedura dello studio e ridiscutere questi negli incontri con i partecipanti, cosicché sia possibile fare correzioni. Nulla fa scemare l'entusiasmo dei clinici a collaborare più che il constatare inutili lungaggini ed il dover compilare schede di registrazione complesse, che fanno perdere tempo. Procedure dello studio semplificate, discusse e concordate prima di iniziare, e verificate in studi pilota, sono importanti per garantire la collaborazione ed il follow-up.

Il protocollo ed il monitoraggio dello studio

Il problema dell'eleggibilità di un paziente per uno studio clinico controllato è risolto dal clinico che ne verifica le condizioni necessarie - test diagnostici, stato clinico, trattamenti precedenti, altre patologie. Il clinico dovrebbe anche assicurarsi che non vi sia un motivo specifico per escludere un paziente, come la preferenza chiara per un trattamento, controindicazioni specifiche o la possibilità di un errore sistematico nella risposta.

Un buon metodo è quello di tenere un registro di tutti i pazienti che soddisfano le condizioni per l'entrata in studio e di registrare accanto ad ogni nome sia il numero progressivo dello studio clinico controllato, sia le condizioni di ingresso od il motivo dell'esclusione del paziente. In questo modo, i ricercatori definiscono il gruppo di pazienti sul cui trattamento verranno fatte le inferenze sulla base dei risultati dello studio, ed inoltre stabiliscono la proporzione di pazienti studiati rispetto a tutti quelli che soddisfano i criteri diagnostici.

In generale, i criteri di ammissione dovrebbero essere meno restrittivi quando il fine sia il confrontare politiche di trattamento e così formulare una decisione pratica sul tipo di terapia che non quando lo studio abbia lo scopo di spiegare, ad esempio, i meccanismi d'azione dei vari trattamenti. Inoltre, in pratica, le condizioni di ingresso saranno di solito tali da assicurare che i pazienti seguano il follow-up in modo completo e preciso. Perciò si può

porre un limite d'età, oppure i pazienti possono essere esclusi a causa di altre patologie, perché la loro aspettativa di vita è minore della durata dello studio (eccetto, ovviamente, il caso in cui la sopravvivenza sia l'evento finale dello studio), perché non si possono muovere o perché le loro famiglie non possono seguirli in modo adeguato. Questi criteri dovrebbero essere definiti in modo semplice, con il consenso dei clinici che partecipano allo studio, ed essere interpretati da questi uniformemente. I ricercatori dovrebbero riconoscere che il limitare la popolazione in studio limita anche la generalizzabilità dei loro risultati.

Oltre a definire i criteri di ingresso, il protocollo dello studio deve fornire dettagli sul disegno, sulla randomizzazione, sui controlli che saranno fatti sulla conduzione della ricerca nonché informazioni su come i dati verranno elaborati. Ma il protocollo è importante anche per un altro motivo, a volte dimenticato, ossia stabilire le ipotesi base a priori dello studio, in modo da distinguerle dalle ipotesi che vengono suggerite una volta noti i risultati dello studio. L'analisi dovrebbe essere completamente programmata già al momento in cui viene deciso il disegno di uno studio clinico controllato. Questa analisi preliminare è quindi inserita nel protocollo ed i programmi per il calcolatore sono preparati in anticipo, come necessario.

Esclusioni specifiche.

I pazienti per i quali uno dei trattamenti in studio è controindicato differiscono per quella controindicazione - e forse anche rispetto ad altri fattori relativi ai risultati - dai pazienti che possono ricevere tutti i trattamenti studiati senza danno. Perciò questi pazienti dovrebbero venir esclusi.

Protocollo

Un'accurata documentazione è importante perché di solito l'articolo su uno studio clinico controllato si scrive mesi o anni dopo che è entrato il primo paziente. Il protocollo costituisce per gli autori uno schema sia delle decisioni prese, sia del tipo di ragionamenti e calcoli che hanno portato a quelle decisioni. Il protocollo, perciò, sarà un punto di riferimento incompleto se ogni cambiamento fatto alla procedura dello studio nel corso della fase di reclutamento dei pazienti e del follow-up non sarà descritto in appendice, con la data, il tipo ed il motivo della modificazione.

La seconda ragione per avere una documentazione completa è il fatto inevitabile che i ricercatori cambiano: i nuovi partecipanti devono avere chiare le procedure dello studio, descritte nel protocollo. Infine, la documentazione dovrebbe essere dettagliata in modo sufficiente da permettere allo statistico di cogliere immediatamente i punti salienti dello studio, al momento di un'analisi intermedia o finale. Altrimenti possono essere trascurati importanti confronti oppure si può riportare un'analisi meno sensibile, perché gli aspetti critici del disegno dello studio sono stati dimenticati.

Il protocollo dovrebbe essere scritto in modo abbastanza chiaro da permettere una ripetizione dello studio in ogni altro luogo e un'analisi analoga. E' utile un promemoria per preparare i protocolli degli studi clinici controllati.

Verifiche sulla conduzione dello studio

Le considerazioni etiche obbligano lo statistico a rivedere periodicamente la massa dei dati provenienti da studi a lungo termine per la presenza di tossicità, di mortalità eccessiva per una causa particolare o di importanti differenze nei trattamenti. Per giustificare correttamente i ripetuti test di significatività, il disegno dello studio dovrebbe specificare il numero di tali analisi provvisorie, e quando esse debbano venir svolte.

La maggior parte degli altri controlli vanno fatti per evidenziare eventuali problemi nella progressione dello studio. Ad esempio una limitazione nel tasso di reclutamento dei pazienti è spesso il primo segno della perdita di entusiasmo per lo studio da parte dei ricercatori, e pertanto spesso del bisogno di un rapporto sullo stato di andamento dello studio. Un cambiamento nel tasso di perdita dei pazienti dallo studio è un'indicazione del fatto che i criteri di accettazione si siano allentati permettendo l'ammissione di pazienti che in precedenza sarebbero stati giudicati non eleggibili, oppure del fatto che vi sia stata una variazione sistematica nella stima della gravità degli effetti collaterali. Altre indicazioni di alterazioni minori nel tipo di pazienti ammessi nello studio sono cambiamenti nella distribuzione delle caratteristiche iniziali del paziente - quale età alla diagnosi, stadio del tumore, istologia - oppure nel livello di risposta a tutti i trattamenti. Cambiamenti nel livello generale di risposta si potrebbero anche spiegare con un diverso criterio nella sua misurazione. Un simile cambiamento solo in un singolo trattamento è parimenti allarmante. Spiegazioni possibili sono un'imprecisione nello schema

di assegnazione al trattamento così da avere un errore sistematico di selezione, un cambiamento nella prescrizione del farmaco (nel numero di somministrazioni, per esempio), oppure una valutazione viziata da errore sistematico perché l'osservatore conosce il trattamento assegnato.

Le fasi pratiche di svolgimento di uno studio clinico controllato sono la parte più affascinante per i clinici; non dovrebbero però essere un periodo di riposo per lo statistico.

Un'eccessiva fretta nell'incominciare l'analisi può compromettere la pianificazione rigorosa e l'esecuzione stessa dello studio; bisogna eseguire la fase importante di controllo dei dati, prima dell'analisi. Nemmeno con una buona competenza nell'uso della statistica si possono rendere corretti dei dati mal raccolti. Ciò che è necessario, in primo luogo, è preparare una verifica dei pazienti persi al follow-up o ritirati dallo studio, dandone le ragioni e valutando i possibili errori sistematici; in secondo luogo, controllare che le osservazioni siano logiche - per esempio entro l'arco dei valori ragionevoli - corrette e coerenti per ogni paziente; inoltre, confrontare le caratteristiche di base dei gruppi di randomizzazione, per verificare che i pazienti in questi gruppi siano simili all'entrata in studio, o per evidenziare quelle possibili differenze che si devono tenere in considerazione durante l'analisi.

Perdita di pazienti

Alcuni pazienti si perdono o vengono esclusi da uno studio per vari motivi, quali il rifiuto del paziente a continuare lo studio, carenze strutturali, perdita al follow-up a causa di un trasferimento, effetti collaterali o tossicità, malattie ricorrenti o morte, constatazione che il paziente non soddisfa i criteri di ingresso, cambiamento di trattamento per cui il paziente aveva ricevuto un trattamento diverso da quello cui era stato randomizzato, scarsa osservanza dei protocolli, per esempio aumento o diminuzione nella dose in rapporto allo stato clinico del paziente. Perciò, oltre a riportare il numero preciso dei pazienti ritirati dallo studio, si dovrebbero specificare in ogni caso le ragioni del ritiro e le loro implicazioni nell'analisi.

I risultati dello studio possono essere seriamente viziati da errori sistematici ignorando il risultato di pazienti che hanno lamentato effetti collaterali, rifiutato di cooperare, mostrato una scarsa collaborazione, o per i quali è stato necessario modificare il regime dei trattamenti. Il più delle volte l'analisi corretta è da farsi tra i gruppi di pazienti così come sono stati definiti dalla randomizzazione e non, successivamente, in base al trattamento. Questo è importante specialmente nel caso di un possibile deliberato cambiamento nel trattamento. Il razionale è quello di un confronto pragmatico, in base alle intenzioni di trattamento - cioè, alle politiche di trattamento. Mentre lo studio viene progettato, i ricercatori dovrebbero elencare le possibili ragioni di esclusione di un paziente, stabilire il miglior modo di analizzare i risultati di ogni paziente al fine di evitare errori sistematici, e scrivere nel protocollo dello studio i metodi di analisi che saranno utilizzati.

Bisogna seguire i pazienti che hanno interrotto la terapia esattamente allo stesso modo di quelli che continuano il trattamento, supposto che il paziente sia d'accordo. I pazienti che non collaborano hanno spesso risultati e prognosi diverse da quelli che collaborano. Le differenze di trattamento possono essere esagerate, diminuite o cambiate da esclusioni non meditate di casi rilevanti.

Confrontabilità

Perché è necessario controllare che i gruppi di trattamento siano effettivamente confrontabili per ciò che riguarda le caratteristiche di base, anche quando il metodo di assegnazione è casuale?

Ci si aspetta che la randomizzazione produca gruppi di trattamento che siano confrontabili per ciò che riguarda le caratteristiche importanti di base, ma ciò non avviene in ogni caso, e perciò il ricercatore deve controllare che vi sia un soddisfacente bilanciamento. Non solo il ricercatore, ma anche il lettore, è rassicurato quando non ci sono ovvie discrepanze iniziali tra gruppi di trattamento. Se non vi è un giusto equilibrio, allora l'autore sa di doverne tener conto nell'analisi, e controllare che non vi sia stato un difetto nella procedura di randomizzazione.

METANALISI

Odds Ratio

Il rischio si può esprimere come percentuale dell'intero campione, vale a dire quando l'evento si esprime 22 volte in un campione di cento persone, diciamo che il rischio è di 0.22 (22%). Lo possiamo esprimere anche come Odds Ratio, cioè come **rapporto tra il numero degli eventi e il numero dei non eventi**.

Nell'esempio, 0.28 (0.22 / 0.78).

Facciamo un esempio pratico. Immaginiamo di trattare con aspirina un postinfarto e di controllare con un campione trattato con placebo la mortalità dopo 1 mese.

	Vivi	Morti
Trattati (gruppo A, n=253)	225	28
Non trattati (gruppo B, n=245)	208	37

Per stabilire con un unico parametro se il trattamento A funziona si usa l'Odds Ratio.

Per calcolare l'Odds Ratio si costruisce una tabella di contingenza 2x2 con i dati dello studio, come sopra.

L'Odds Ratio è definito come:

	Vivi	Morti
Casi	a	b
Controlli	c	d

$$OR = (a * d) / (c * b)$$

$$OR = (225 * 37) / (208 * 28) = 1.42.$$

Scambiando i due gruppi si ottiene un risultato che possiamo definire speculare:

$$OR = (208 * 28) / (225 * 37) = 0.7$$

Quale valore di OR prendere in considerazione: 1.42 o 0.7.

Il rischio di morte per il gruppo A è $28/253 = 11\%$ e per il gruppo B è $37/245 = 15\%$.

Quindi essendo il rischio di morte per il gruppo A minore di B, ovviamente l'OR sarà 1.42 perché essendo l'OR considerato come un'approssimazione del rischio relativo esprime che il gruppo B ha un rischio di 1.42 volte superiore rispetto al gruppo A.

Una volta calcolato l'OR si stabilisce dove si trova il valore di OR rispetto a 1, che rappresenta il valore nullo o di non effetto. Per convenzione si stabilisce che se si trova a destra della linea di non effetto significa che il trattamento è migliore, se a sinistra significa che è peggiore dei controlli. Inoltre occorre calcolare gli intervalli di confidenza.

La formula per il calcolo degli intervalli di confidenza è la seguente:

Prima si calcola l'errore standard di OR che sarà:

$$ES(\log_e OR) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$$

Poi

$$Y = \log_e OR - (1.96 * ES(\log_e OR))$$

$$Z = \log_e OR + (1.96 * ES(\log_e OR))$$

$$IC\ 95\% = e^Y \text{ e } e^Z \text{ dove } e \text{ è la costante di Nepero (2.718)}$$

Nel nostro esempio abbiamo:

$$ES(\log_e OR) = \sqrt{\frac{1}{208} + \frac{1}{37} + \frac{1}{225} + \frac{1}{28}} = 0.268$$

$$\log_e OR = 0.351$$

$$Y = 0.351 - (1.96 * 0.268) = -0.174$$

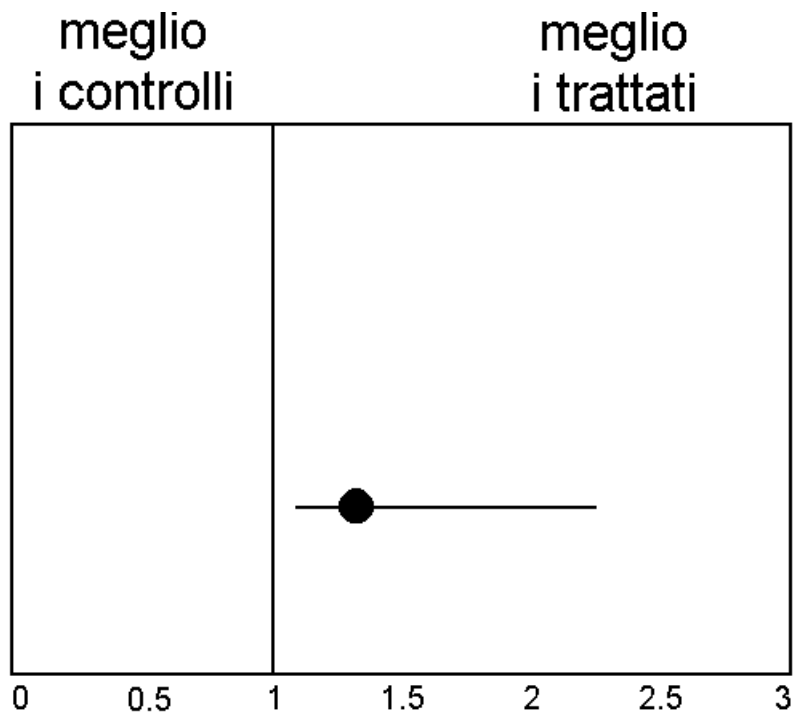
$$Z = 0.351 * (1.96 * 0.268) = 0.876$$

$$e^{-0.174} = 1.2$$

$$e^{0.876} = 2.4$$

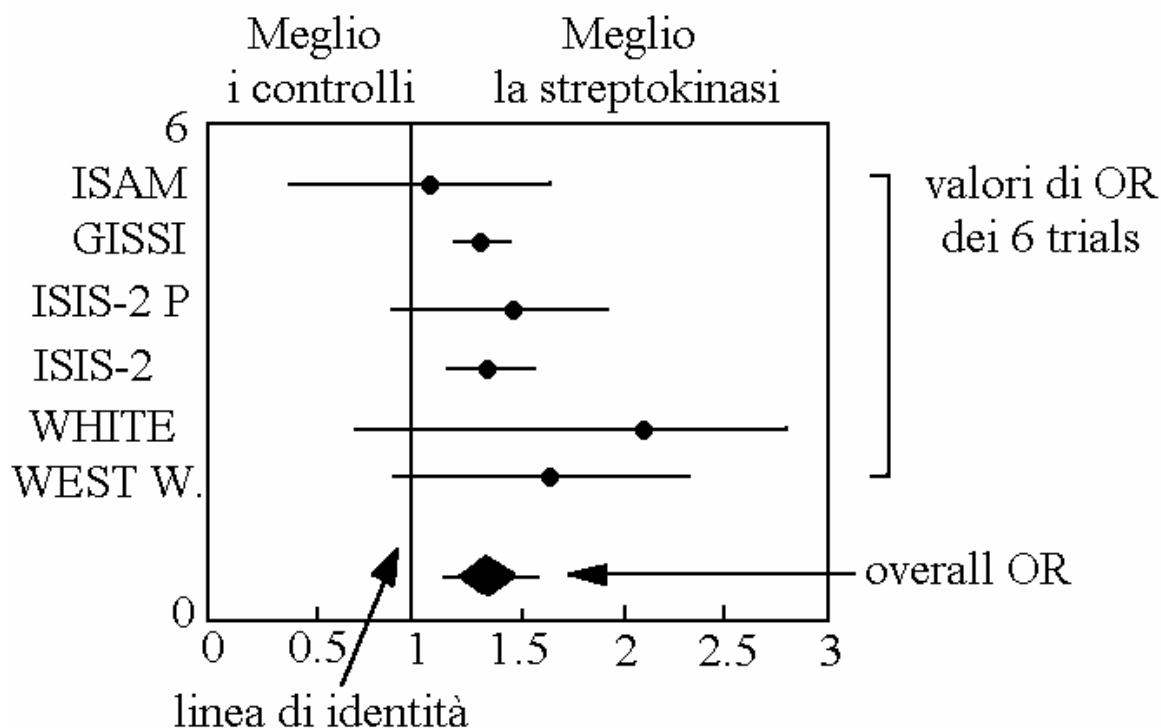
Quindi OR 1.42 (IC 95% 1.2-2.4)

Quindi il farmaco è a dx nel grafico e non contiene 1, per cui è superiore al placebo.



I diagrammi di odds-ratio presentati nelle metanalisi di qualità intendono semplicemente rappresentare graficamente in modo comprensibile ed accattivante risultati spesso complicati.

Finora si è parlato di singolo studio, ma come fare a estrapolare i dati presi da diversi studi e tradurli in un singolo OR per stabilire se un certo tipo di trattamento è efficace o no? In altre parole noi abbiamo diversi studi con diversi OR, alcuni a favore del trattamento e altri contro. Come tradurli in un singolo OR? Questo singolo OR viene chiamato **Overall OR**.



Altro esempio: si tratta di una metanalisi di Midgette (1990) sulla efficacia della streptochinasi nell'infarto. Gli studi inclusi nella metanalisi sono sei:

- ISAM : negativo
- GISSI : positivo
- ISIS-2 pilota : negativo
- ISIS-2 : positivo
- WHITE : negativo
- WEST.WASHINGTON : negativo

Risultato della Metanalisi: Il rombo è situato a destra con intervalli di confidenza a destra che non contengono 1 cioè la linea di nessun effetto. Conclusione. La metanalisi è positiva.

La metanalisi dell'OR si basa sul calcolo dell'Overall OR a partire dai dati dei singoli studi.

Il calcolo è si basa sulla formula di Mantel-Haenzsel.

$$\text{Overall OR} = \frac{\sum \frac{a_i * d_i}{n_i}}{\sum \frac{b_i * c_i}{n_i}}$$

Dove a_i, b_i, c_i, d_i sono le frequenze di ogni iesima tabella 2 x 2, $n_i = a_i + b_i + c_i + d_i$ e \sum è la sommatoria da $i = 1$ a $i = K$ per le K tabelle necessarie per il controllo all'interno dei singoli strati dei potenziali fattori di confondimento.

Alternativamente si può usare il metodo logit per ottenere una stima "pooled" dell'odds ratio (OR_L) e quindi derivare gli intervalli di confidenza per l'odds ratio in modo simile a quello utilizzato per le tabelle 2 x 2.

La stima logit combinata (OR_L) è definita da:

$$\log_e OR_L = \sum w_i \log_e OR_i / \sum w_i$$

dove $OR_i = a_i * d_i / b_i * c_i$ è l'odds ratio nelle i -esima tabella e

$$w_i = \frac{1}{1/a_i + 1/b_i + 1/c_i + 1/d_i}$$

l'errore standard di $\log_e OR_L$ è dato da

$$ES(\log_e OR_L) = \frac{1}{\sqrt{w}}$$

dove $w = \sum w_i$

Gli intervalli di confidenza di OR_L con valori $100(1-\alpha)\%$ si trovano prima calcolando le due quantità

$$M = \log_e OR_L - [N_{1-\alpha/2} * ES(\log_e OR_L)]$$

$$N = \log_e OR_L + [N_{1-\alpha/2} * ES(\log_e OR_L)]$$

dove $N_{1-\alpha/2}$ è il valore appropriato derivato dalla distribuzione normale standardizzata per il percentile $100(1-\alpha/2)$ che si trova nelle tabelle di alfa.

Gli intervalli di confidenza per i valori di OR si calcolano elevando all'esponenziale M e N: e^M ed e^N .

Esempio:

Metanalisi della esposizione al fumo e cancro polmonare.

Studi	Casi di Cancro Polmonare		Controlli		Odds Ratio
	Esposti a	Non Esposti b	Esposti c	Non Esposti d	
1	14	8	61	72	2.07
2	33	8	164	32	0.80
3	13	11	15	10	0.79
4	91	43	254	148	1.23

La stima logit combinata degli odds ratio (OR_L) dei quattro studi è risultata pari a 1.19. L'errore standard di $\log_e(OR_L)$ è 0.1693, che utilizzato per gli intervalli di confidenza al 95% produce i seguenti risultati:

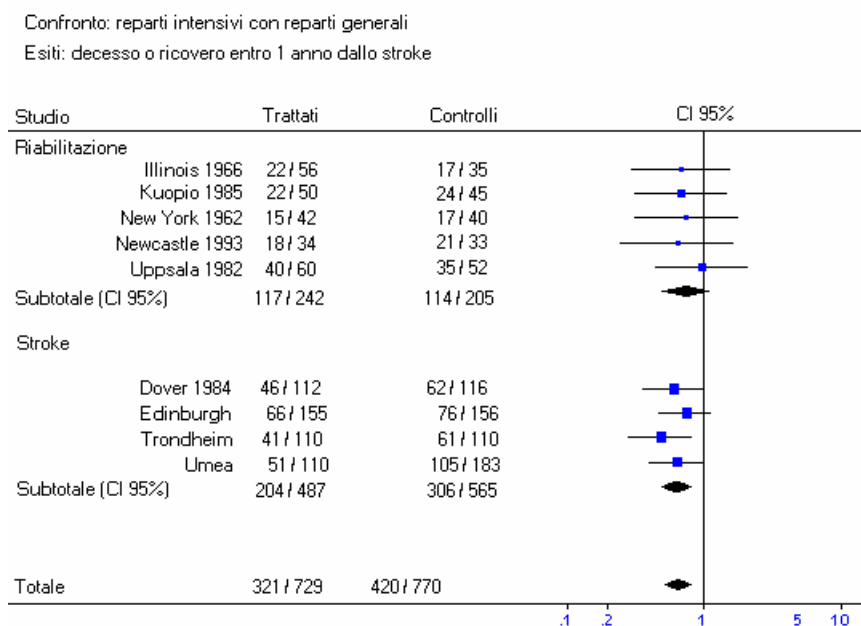
$$M = \log_e 1.19 - (1.96 * 0.1693) = - 0.1579$$

$$N = \log_e 1.19 + (1.96 * 0.1693) = 0.5058$$

Gli intervalli di confidenza al 95% della stima dell'odds ratio erano pertanto compresi tra $e^{-0.1579}$ ed $e^{0.5058}$ ossia 0.85 e 1.66.

Utilizzando la formula di Mantel-Haenzsel, la stima combinata degli odds ratio risulta 1.19, con intervalli di confidenza compresi tra 0.86 e 1.66; cioè lo stesso risultato

La figura sottostante mostra un altro esempio. Il diagramma si riferisce al confronto tra sezioni ospedaliere dedicate agli ammalati di stroke e rimanenti reparti. Per ogni singolo trial il risultato dell'odds ratio è rappresentato con un quadrato. La linea orizzontale attraverso esso rappresenta l'intervallo di confidenza per quel risultato. Il risultato della meta-analisi è rappresentato da un rombo (Overall OR), il cui spessore rappresenta l'intervallo di confidenza. L'asse orizzontale mostra l'odds ratio. La linea verticale rappresenta un odds ratio di 1 ed è conosciuta come *linea di nessun effetto*.



Per poter estrarre informazioni utili da un diagramma di odds ratio è necessario stabilire:

- La natura dell'intervento che si sta esaminando, che è specificata dal titolo
- I risultati che vengono considerati, scritti subito sotto il titolo
- Se ogni risultato è positivo o negativo
- Se il risultato totale è a destra o a sinistra della *linea di nessun effetto*
- Se la linea dell'intervallo di confidenza attraversa la linea di nessun effetto
- Se il risultato complessivo (rombo nero in basso) è chiaramente comprensibile

Se un determinato intervento consegue un risultato negativo (decesso, ad esempio), l'odds-ratio sarà a sinistra della *linea di nessun effetto*, e si avvicinerà alla percentuale del gruppo di trattamento che andrà incontro a decesso, paragonata con quella del gruppo di controllo. Parimenti, se il risultato del trattamento sarà invece positivo, l'odds-ratio sarà a destra della linea verticale di 1.

Se la linea orizzontale dell'intervallo di confidenza (CI) attraversa la *linea di nessun effetto*, allora il risultato (rappresentato da un rombo nero se riguarda una meta-analisi e da un quadretto blu se riguarda un RCT) non dimostra un'efficacia chiara. Gli intervalli di confidenza usati nelle metanalisi sono del 95% o del 99%, e i diagrammi possono essere adattati all'uno o all'altro valore. L'intervallo di confidenza rappresenta il valore in cui, con una probabilità del 95%, si trova il risultato del trial o della meta-analisi, risultato che è estrapolato dal complesso della popolazione. In altri termini, significa che in 95 trial su 100 possiamo essere certi che il risultato sarà compreso nella linea dell'intervallo di confidenza. Se questa linea attraversa la *linea di nessun effetto*, dal momento che possiamo essere certi solamente al 95% che il risultato sia su questa linea, è possibile che un risultato apparentemente positivo sia in realtà negativo. In tal caso la conclusione non può essere attendibile.

Il risultato leggibile sul diagramma deve comunque essere integrato dal testo e dalle implicazioni date nell'intero studio: la rappresentazione grafica non può e non deve sostituire l'analisi accurata di tutta la review.

In sintesi:

- Un'odds ratio a sinistra della linea verticale significa risultato inferiore
- Un'odds ratio a destra della linea verticale significa risultato superiore
- Un risultato positivo per un esito clinico negativo è un'odds ratio a sinistra della linea verticale
- Un risultato positivo per un esito clinico positivo è un'odds ratio a destra della linea verticale

WEIGHTED MEANS DIFFERENCE (WMD)

Alcuni diagrammi di meta-analisi includono nei risultati anche la *weighted means difference*, che è una tecnica statistica differente, usata per valutare differenti tipi di esiti. Questi esiti appartengono ad un genere "dicotimico", in cui un risultato è il contrario dell'altro: vivo/morto, accadimento/non accadimento, guarigione/non guarigione. All'interno dei due estremi non esiste altro risultato. Questi esiti possono essere confrontati mediante un'odds ratio, procedimento che non potrebbe essere attuato con risultati rappresentati su una linea continua (perdita di peso, aumento di pressione, statura, giorni di ricovero, e così via). Nel rappresentare la WMD, la *linea di nessun effetto* cade a 0 invece che a 1, e nelle CDSR la WMD è rappresentata da quadretti verdi. L'interpretazione è comunque identica a quella dei risultati dell'odds ratio: ad esempio, un risultato che stia a sinistra della *linea di nessun effetto* (cioè una WMD inferiore a 0) significa che il risultato è meno probabile che venga ottenuto nel gruppo trattato rispetto al gruppo di controllo.

STATISTICA EPIDEMIOLOGICA

PREVALENZA E INCIDENZA

I dati possono essere espressi in forma di frequenza assoluta e relativa.

La **frequenza assoluta** indica semplicemente la consistenza di un determinato fenomeno. Ad esempio, la quantità di persone con pressione arteriosa < 140 mmHg; il limite di questo valore risiede nell'incapacità di definire l'esatta dimensione del fenomeno nella popolazione, dal momento che il dato globale non è noto: 100 pazienti con ipertensione assumono diverso significato in una popolazione di 1000, 10.000 o 100.000 persone.

La **frequenza relativa** esprime invece il rapporto tra un certo numero di eventi rispetto alla popolazione in studio. 10 pazienti ipertesi su una popolazione di 100 significa una frequenza relativa del 10%.

La **prevalenza** è utile quando si vuole quantificare l'entità di un determinato momento senza particolare interesse per una valutazione epidemiologica globale della situazione. La prevalenza non permette di pronosticare il futuro, perché non registra la sequenza con cui si verificano gli eventi e perché, di fatto, identifica un numero di casi inferiore a quelli calcolati.

L'**incidenza** propone invece un'immagine più realistica del fenomeno, poiché dovrebbe evidenziare tutti i nuovi casi di malattia e con essi la velocità di crescita.

Prevalenza e incidenza appaiono strettamente correlate: in malattie croniche, la prevalenza tende ad aumentare in rapporto all'aumento dell'incidenza e alla riduzione del numero dei soggetti che muoiono o guariscono, mentre, viceversa, tende a diminuire con la diminuzione dell'incidenza e l'aumento dei soggetti che guariscono o muoiono.

In definitiva, i dati di prevalenza e incidenza nel loro insieme servono a tre scopi:

- descrivere i fenomeni come si manifestano nella realtà
- effettuare comparazioni
- formulare previsioni sulla situazione futura

TASSI E QUOZIENTI

In statistica, i tassi sono utilizzati per descrivere le variazioni di un determinato parametro all'interno di una popolazione presa in esame. Il tasso consiste quindi nel rapporto tra il numero degli eventi relativi ad una popolazione e le dimensioni di quest'ultima, moltiplicato per una costante di tempo (mese, anno, quinquennio, decennio). Il fattore tempo, cioè l'intervallo cui si riferisce l'analisi (generalmente un anno solare) è di fondamentale importanza nel concetto di tasso, con l'unica eccezione del tasso di prevalenza. Quest'ultimo è in realtà una frequenza che indica la parte di popolazione che, in un determinato momento, è affetta da una particolare patologia:

Tasso di Prevalenza = (malati / popolazione) * k

Il tasso di incidenza di una malattia indica invece il rapporto tra i nuovi casi patologici osservati nell'unità di tempo e la popolazione inizialmente esposta al rischio, cioè i soggetti non ammalati all'inizio del periodo di osservazione:

Tasso di Incidenza = (nuovi ammalati / popolazione) * k

Il rapporto tra due tassi di incidenza di malattia in due sottopopolazioni caratterizzate da un elemento distintivo (ad esempio presenza/assenza di un fattore di rischio per la malattia studiata) costituisce il cosiddetto rischio relativo (in inglese, Risk Ratio o RR), un indice di misura dell'associazione tra lo svilupparsi della malattia ed il fattore di rischio:

RR = incidenza esposti al fattore di rischio / incidenza non esposti

Immaginiamo di aver seguito per 10 anni 100 soggetti epatopatici cronici e 100 non epatopatici e di aver osservato 7 cirrosi nel primo gruppo e 2 nel secondo. L'incidenza della patologia risulta rispettivamente di 0.07 e 0.02 ed il RR di 3.5 (0.07/0.02). Si può concludere che un'epatopatia cronica espone ad un rischio 3.5 volte maggiore di sviluppare una cirrosi epatica, rispetto al rischio posseduto da una popolazione sana. Organizzando i dati in una tabella 2 x 2, nel caso di una patologia rara, il rischio relativo è ottenibile dal prodotto dei rapporti dei valori in diagonale:

$$RR = a / b * d / c$$

	Malati	Sani
Esposti	a	c
Non esposti	b	d

Analogamente al tasso di incidenza è possibile calcolare, utilizzando opportuni numeratori e denominatori, una serie di tassi specifici in grado di fornire indicazioni molteplici.

$$\text{Tasso di morbilità} = (\text{giorni di malattia} / \text{popolazione a rischio}) * k$$

dove per popolazione a rischio si intende l'insieme dei soggetti suscettibili di sviluppare la patologia oggetto di studio (ad es. l'insieme dei ricoverati se l'indagine viene effettuata per quanto riguarda le infezioni ospedaliere). Indica la durata dello stato patologico ed è generalmente più elevato per le malattie croniche, o, in certi periodi, per le cosiddette patologie stagionali.

$$\text{Tasso di letalità} = (\text{morti da causa specifica} / \text{casi diagnosticati}) * k$$

indica la gravità della malattia ed è diverso dal tasso di mortalità

$$\text{Tasso di mortalità} = (\text{morti da causa specifica} / \text{popolazione}) * k$$

La mortalità infatti dipende, oltre che dalla gravità della patologia esaminata, anche dalla prevalenza di tale patologia nella popolazione presa in esame.

NUMERI INDICE

Un'evoluzione temporale di un fenomeno può essere descritta ricorrendo ai rapporti incrementali che corrispondono alla variazione percentuale nell'unità di tempo:

$$p = (\Delta Q / Q_0) * 100$$

dove p è la percentuale; ΔQ è la variazione quantitativa tra due osservazioni consecutive nel tempo ($Q_1 - Q_0$); Q_0 è la quantità (o frequenza) al tempo basale dell'intervallo considerato.

Esempio: Immaginiamo di voler valutare l'effetto di alcuni interventi di profilassi sull'incidenza di infezioni ospedaliere in un reparto chirurgico. Teniamo sotto osservazione il reparto per tre mesi (Q_1) osservando 20 infezioni, dopodiché instauriamo il programma di profilassi e continuiamo il monitoraggio per un altro trimestre osservando solo 5 infezioni.

$$\text{Il } \Delta Q \text{ è pari a } 15 \text{ (} 20 - 5 \text{) per un valore di } p = (15 / 20) * 100 = - 75\%$$

L'incremento percentuale è in realtà un dato che può assumere valori negativi e che si rapporta sempre ad un valore di confronto precedente (Q_0). Un altro modo per interpretare le variazioni temporali di un evento è costituito dai numeri indice che corrispondono al rapporto tra due dati (quantità, frequenze) separati da un periodo di tempo arbitrariamente scelto come unitario (mese, anno).

Numero indice = F_1 / F_0

È rappresentato da una osservazione finale (F_1) rapportata ad un dato iniziale (F_0) ed è un valore assoluto espresso senza alcuna unità di misura. In pratica i numeri indice vengono espressi sotto forma di frequenze relative o percentuali in cui la grandezza usata al denominatore come valore di riferimento è convenzionalmente posta rispettivamente uguale a 1 o a 100.

Riferendoci all'esempio il numero indice per le infezioni rilevate dopo l'introduzione delle misure profilattiche è 0.25 (25%).

NNT e NNH

	Trial A		Trial B		Trial C	
	Trattati	Controlli	Trattati	Controlli	Trattati	Controlli
Pazienti (n)	100	100				
Incidenza (i)	5	30				
RA (i/n)	0.05 (a)	0.30 (b)	0.14 (a)	0.8 (b)	0.00017 (a)	0.001 (b)
RR (a/b)	0.17 (c)		0.17 (c)		0.17 (c)	
RRR ([b-a]/b oppure 1-c)	0.83 (d)		0.83 (d)		0.83 (d)	
RRA b-a	0.25 (e)		0.66 (e)		0.00083 (e)	
NNT 1/e	4		1.5		1204	

RR: Rischio Relativo

RRR: Riduzione del Rischio Relativo

RRA: Riduzione del Rischio Assoluto

NNT: Numero Necessario da Trattare

Nel trial A risulta un NNT di 4. Significa che devo trattare 4 pazienti per evitare una emicrania in un certo periodo di tempo, che di solito viene espressa dalla durata del trial. Osserviamo l'esempio sopra. A parità di RR e RRR per i 3 trials abbiamo 3 diversi NNT. Come si vede il NNT ha il vantaggio della facilità del calcolo e dell'immediatezza; può inoltre essere usato per quantificare gli effetti collaterali di un farmaco: in questo caso si parla più propriamente di NNH (number needed to harm: numero necessario per avere un danno).

Più basso è il NNT e più efficace è la terapia.

Bisogna però distinguere i trials terapeutici, il cui fine è verificare l'efficacia di un farmaco (ad es. antibiotico versus placebo, in cui il NNT sarà molto basso, all'incirca vicino a 1), dai trials il cui fine è la prevenzione (ad es. aspirina + streptochinasi nella prevenzione a 5 settimane della mortalità vascolare dopo IMA, in cui un NNT tra 20 e 40 è considerato clinicamente efficace).

Come si può notare NNT è contesto-specifico, dipende dal fine del trial.

Limiti del NNT

- IL NNT è espresso come singolo numero stimato teorico. Tuttavia il valore vero di NNT può essere più alto o più basso rispetto al teorico. Il 95% dell'Intervallo di Confidenza (CI) è utile a questo scopo perché esso da indicazione che 19 volte su 20 il valore vero di NNT cade in uno specifico range. Un NNT con un CI infinito è solo un punto stimato; esso include la possibilità di nessun beneficio o danno, potendosi discostare enormemente dalla situazione reale.
- E' sbagliato paragonare il NNT tra condizioni diverse, particolarmente quando i risultati differiscono. Ad esempio, un NNT di 30 per prevenire una trombosi venosa deve essere valutato diversamente da un NNT di 30 per la prevenzione dello stroke o per la prevenzione della morte.
- IL NNT non è una quantità fissa, in quanto dipende non solo dalla natura del trattamento ma anche dal rischio basale. Poiché il rischio non è lo stesso per tutti i pazienti, un NNT, che è fornito dalla letteratura, deve essere aggiustato per compensare il rischio di base per ogni singolo paziente. In altre parole, se il rischio è la morte o una condizione grave, e tale rischio è elevato, il medico dovrà decidere se trattare o no il

suo paziente. (Il NNT di uno studio che valutava la riduzione dello stroke con l'uso di un farmaco anti-ipertensivo era 12.5, su una popolazione con un determinato rischio cardiovascolare. Se ho un paziente con un rischio basale della metà rispetto al rischio basale dei pazienti dello studio, allora devo correggere NNT adattandolo al nostro paziente. Vale a dire $12.5 / 0.5 = 25$. Se gli intervalli di confidenza erano 11.4-13.9 dovrò correggere anche loro adattandoli al nostro paziente. In altre parole $11.4/0.5$ e $13.9/0.5$ cioè 22.8-27.7).

- Un NNT è sempre calcolato su uno specifico periodo di tempo: un NNT ricavato da un trial durato un anno non può sic et simpliciter essere adottato su periodi di tempo minori o maggiori.

Trattamento

Un importante aspetto da tenere in mente nel considerare un NNT favorevole è la presenza di eventuali effetti collaterali ed il costo della terapia.

Confrontando, ad esempio, il sumatriptan sottocute (NNT 2) con quello orale (NNT 2.6) e con placebo per la cessazione dell'emicrania a 2 ore, si nota che a causa dell'alto costo della forma sottocute è preferibile usare la forma orale in quanto la differenza tra i due NNT è minima. Non solo, ma se uno studio con aspirina + metoclopramide ha dimostrato un NNT di 3, la differenza minima con gli altri due studi e il suo più basso costo porterà il clinico a prendere in considerazione questa terapia come prima scelta.

Profilassi

Paragonare il rischio di un evento senza e con la profilassi è una questione più complessa rispetto al trattamento: come per la terapia, occorre considerare non solo il NNT, ma anche i costi; inoltre deve essere tenuto presente, nella decisione, anche il fine della profilassi. Al termine della valutazione, il medico o il politico decideranno le scelte da intraprendere.

Facciamo un esempio concreto. Uno studio apparso sugli *Annals of Internal Medicine* (15.08.95;123:241-9) sulla profilassi del sanguinamento gastroenterico con misoprostolo ha dimostrato un NNT di 263 a 6 mesi. Questo ci porta a non considerare la profilassi. Ma analizzando i dati e considerando i soggetti a rischio (cardiopatici, precedente ulcera peptica, età > 75 anni e precedente sanguinamento) gli NNT a 1 anno diventano:

Pazienti	Tutte le età	Età > 65	Età > 75
Tutti	132	76	105
Precedente evento cardiovascolare	68	53	58
Precedente ulcera peptica	23	16	11
Precedente sanguinamento GI	17	14	7

Un altro esempio. L'uso di statine nella prevenzione secondaria di un evento cardiovascolare in un individuo con 6% di rischio ha un NNT di 13 a 5 anni. Devo trattare per 5 anni 13 persone per prevenire un evento cardiovascolare. Ma quale sarà il costo?

Considerando che il costo della simvastatina nel 1998 è di 27.600 lire per una scatola di 10 compresse, risulta una spesa di 65 milioni in 5 anni per evitare un evento cardiovascolare. Vale la pena?

NNH

Il concetto richiamato dal NNT può essere usato per le reazioni avverse (NNH). Quando l'incidenza degli effetti avversi è bassa è facile che gli Intervalli di Confidenza non siano disponibili; sarà utilizzabile quindi solamente il punto stimato del danno.

COME CALCOLARE IL NNT E L'INTERVALLO DI CONFIDENZA

Non esiste un solo modo per ottenere il valore del NNT. Le procedure possibili prevedono:

Calcolo del NNT dai dati crudi secondo la seguente formula

	Trattamento attivo	Controlli
Totale	Ta	Tc
Migliorati	Ia	Ic

$$\text{NNT} = \frac{1}{\left(\frac{Ia}{Ta}\right) - \left(\frac{Ic}{Tc}\right)}$$

Calcolo dell'Intervallo di Confidenza (CI)

Il metodo Pukka consiste nell'invertire e scambiare i limiti di un intervallo di confidenza del 95% per la riduzione del rischio assoluto (RRA). La formula di calcolo che ne deriva è la seguente:

$$\text{CI} = \frac{1}{\left[\left(\frac{Ia}{Ta}\right) - \left(\frac{Ic}{Tc}\right)\right]} \pm 1.96 * \sqrt{\frac{\left(\frac{Ia}{Ta}\right) * \left[1 - \left(\frac{Ia}{Ta}\right)\right]}{Ta} + \frac{\left(\frac{Ic}{Tc}\right) * \left[1 - \left(\frac{Ic}{Tc}\right)\right]}{Tc}}$$

Per la profilassi il NNT è dato dall'equazione: $\frac{1}{Ic - Ia}$, dove Ic ed Ia sono le proporzioni di coloro che beneficiano della profilassi nel gruppo di controllo e nel gruppo trattato.

Per il trattamento l'equazione è lievemente modificata: $\frac{1}{Ia - Ic}$

Il NNT per i trattamenti dovrebbe essere il più piccolo possibile, possibilmente non superiore a 4. La triplice terapia per la eradicazione dell'*Helicobacter Pylori* ha un NNT di 1.2.

Nella profilassi invece il NNT è alto in quanto pochi pazienti beneficeranno della profilassi. L'uso di aspirina per la prevenzione di una morte cardiovascolare a 5 settimane ha un NNT di 40.

Riduzione del Rischio Assoluto (RRA)

L'RRA si calcola dalla differenza tra la percentuale dell'evento nel gruppo di controllo e la percentuale dell'evento nel gruppo trattato. È il denominatore nel calcolo del NNT.

I trials seri mettono sempre questo dato, per cui è semplice ricavare NNT come = 1/RRA

NNT e ODDS RATIO

Il NNT dall'Odds Ratio si calcola dalla seguente Tabella.

		Treatment OR									Prevention OR								
		0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	1.5	2	2.5	3	3.5	4	4.5	5	10
	0.05	41	46	52	59	69	83	104	139	209	43	22	15	12	9	8	7	6	3
Control	0.1	21	24	27	31	36	43	54	73	110	23	12	9	7	6	5	4	4	2
Event	0.2	11	13	14	17	20	24	30	40	61	14	8	5	4	4	3	3	3	2
Rate	0.3	8	9	10	12	14	18	22	30	46	11	6	5	4	3	3	3	3	2
(CER)	0.4	7	8	9	10	12	15	19	26	40	10	6	4	4	3	3	3	3	2
	0.5	6	7	8	9	11	14	18	25	38	10	6	5	4	4	3	3	3	2
	0.7	6	7	9	10	13	16	20	28	44	13	8	7	6	5	5	5	5	4
	0.9	12	15	18	22	27	34	46	64	101	32	21	17	16	14	14	13	13	11

L'Odds Ratio si trova sulla linea in alto e la frequenza dell'evento dei controlli sulla linea a sn. Gli NNT si trovano nelle caselle. Se abbiamo un Odds Ratio di 0.6 e un CER di 0.5 troviamo il NNT nella casella in cui i due valori si incrociano (NNT = 8).

L'Odds Ratio deve essere interpretato con cautela quando la frequenza dell'evento è alta, come nei trattamenti, e può sovrastimare il beneficio di un effetto quando la frequenza dell'evento sia oltre il 10%. In questo caso l'Odds Ratio va sostituito dalla Riduzione del Rischio Relativo perché esso da informazioni migliori in situazioni dove la frequenza dell'evento è alta.

Oppure lo si può calcolare ricorrendo alle formule seguenti:

La formula per calcolare il NNT per terapie preventive è:
$$NNT = \frac{\{1 - [CER * (1 - OR)]\}}{[(1 - CER) * CER * (1 - OR)]}$$

La formula per calcolare invece il NNT per interventi terapeutici è:
$$NNT = \frac{[CER(OR - 1) + 1]}{[CER(OR - 1) * (1 - CER)]}$$

Se usiamo l'esempio del farmaco per la terapia dell'emicrania esposto sopra (Trial A) abbiamo:

$$NNT = \frac{[0.30 * (0.12 - 1) + 1]}{[0.30 * (0.12 - 1) * (1 - 0.30)]} = \frac{0.30 * (-0.88 + 1)}{0.30 * -0.88 * 0.70} = \frac{0.736}{-0.1848} = 3.98 \text{ cioè } 4$$

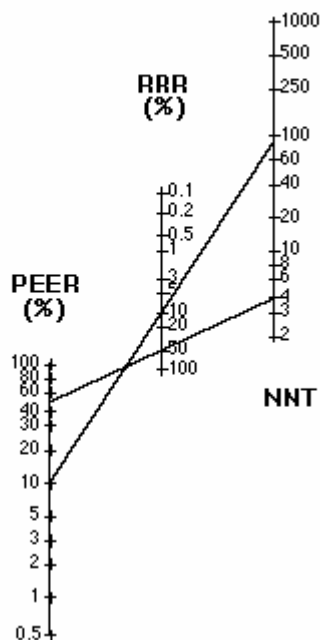
4. La riduzione del Rischio Relativo (RRR)

L'Odds Ratio deve essere interpretato con cautela in caso di eventi che occorrono comunemente come in caso di trattamenti per malattie. In questi casi l'Odds Ratio può essere sovrastimato, ed è utile ricorrere alla riduzione del Rischio relativo.

IL NNT può essere calcolato da un Nomogramma ideato da Chatelier.

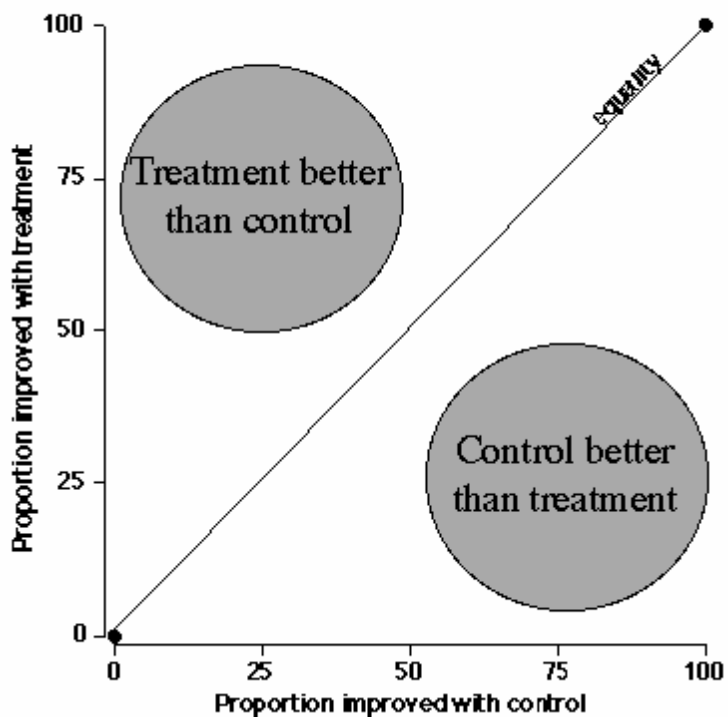
NOMOGRAM FOR NNTs

(From [5])



L'Abbé plots: si tratta di un semplice grafico per rappresentare i dati presi dai trials.

L'Abbé plot for treatment



Per il trattamento il trial in cui l'intervento era migliore rispetto al gruppo di controllo sarà in alto a sinistra compreso tra l'asse y e la linea mediana. Se il trattamento non era meglio del gruppo di controllo allora il punto cadrà sulla linea di uguaglianza, e se il controllo era meglio del trattamento allora il punto sarà in basso a destra del grafico tra l'asse della X e la linea di uguaglianza.

Per la profilassi il tutto andrà rivisto. Poiché la profilassi riduce il numero di cattivi eventi - come morte dopo IMA con l'uso di aspirina - ci aspetteremo una più piccola percentuale di eventi avversi con il trattamento che nel controllo. Così se il trattamento è meglio che il controllo il trial sarà tra la X e la linea di uguaglianza. Questo grafico dà un veloce giudizio sui trials. Se i punti sono in una nuvola omogenea ci dice che i dati sono omogenei. Se i punti sono sparsi nel grafico e specialmente se essi incrociano la linea di uguaglianza allora ci dirà che i dati sono eterogenei.

TABELLA degli NNT

Condizione	Trattamento	Durata trattamento	Controlli	Risultato	Odds Ratio	NNT (95% CI)	Rif.
AIDS	Indinavir (tripla terapia con nucleosidi)	38 settimane	Doppia terapia (nucleosidi e placebo)	Primo risultato clinico (AIDS o decesso)		19 (12 - 50)	
Angina	Bisoprololo	8 settimane	Nifedipina	Completa assenza di sintomi		2.7 (2.1 - 3.9)	
Angina	Bisoprololo	4 - 6 settimane	Placebo	Angina da sforzo		2.8 (1.9 - 5.0)	
Angina	Isosorbide dinitrato	4 - 6 settimane	Placebo	Angina da sforzo		5.0 (2.8 - 21)	
Arteriopatia periferica	Naftidrofuryl	3 o 6 mesi	Placebo	Miglioramento della claudicatio del 50% dopo 1 anno	1.5 (1.2 - 2.0)	10.3 (6.3 - 29)*	8
Arteriopatia periferica	Naftidrofuryl	3 o 6 mesi	Placebo	Prevenzione di eventi critici cardiaci a 1 anno	0.6 (0.4 - 0.96)	24 (13 - 266)*	8
Artrite	Glucosamina	3 - 8 settimane	Placebo	miglioramento		5 (3.5 - 8.9)	
Asma infantile	Assistenza infermieristica	4 settimane	Solita assistenza	Ricovero ripetuto		6.1 (3.8 - 15)	29
Asma infantile	Budesonide e formoterolo	1 anno	Budesonide	Guarigione o grave peggioramento per un anno		11 (6.6 - 34)	
Colite ulcerosa	Nicotina transdermica	4-6 settimane	Placebo	Miglioramento		3.7 (2.4 - 8.4)	
Demenza	Ginkgo	1 anno	Placebo	4 punti in più sulla scala ADAS		7.9 (4.2 - 67)	
Depressione infantile	Antidepressivi	Non stabilita	Placebo	Miglioramento	1.1 (0.5 - 2.2)	Non efficaci	9
Disfunzione erettile	Alprostadil transuretrale	Oltre 3 mesi	Placebo	Erezione raggiunta		2.3 (2.1 - 2.6)	
Distorsione acuta	FANS topici			Riduzione del dolore al 50%		2+	
Dolore postoperatorio	Ibuprofen 400 mg			Miglioramento del 50%		2	
Dolore postoperatorio	Codeina 60 mg			Miglioramento del 50%		18	
Dolore post-operatorio	Paracetamolo 1 g	Una dose	Placebo	Miglioramento al 50%		3.6 (3.0 - 4.4)	5
Emicrania	Sumatriptan sottocute	Una dose	Placebo	Scomparsa della cefalea in 2 ore		2.0 (1.8 - 2.2)*	3
Emicrania	Sumatriptan orale	Una dose	Placebo	Scomparsa della cefalea in 2 ore		2.6 (2.3 - 3.2)*	3
Emorragie da FANS	Misoprostolo	6 mesi	Placebo	Prevenzione di ogni complicanza GI	0.6 (0.4 - 0.85)	166 (97 - 578)*	16
Esofagite	Omeprazolo	8 settimane	Ranitidina	Guarigione endoscopica		3.3	
Esofagite	Omeprazolo	1 anno	Ranitidina	Guarigione		2.8	
Helicobacter pylori	Tripla terapia			eradicazione		1.1	36-3
Herpes zoster	Acyclovir	5-10 giorni	Placebo	Prevenzione di nevrite posterpetica a 6 mesi	0.7 (0.5 - 1.1)	Non efficace	17
IMA	ACE inibitori [studio AIRE]			Morte entro 6 mesi		18	11-1
IMA	ASA e streptochinasi	Infusione e.v. di streptochinasi per 1 ora + ASA per os per un mese	Nessuna cura	Mortalità a 5 settimane: prevenzione di un decesso		20*	15
IMA e diabete	Insulina	3 anni	Solite cure	Mortalità ad 1 anno		14 (7.3 - 164)	
Influenza	Vaccinazione			Assenza di influenza		23	11-5
Ipertensione nell'anziano	Farmaci	Almeno 1 anno	Nessuna terapia	Prevenzione di eventi cardiovascolari a 5		18 (14 - 25)	14

				anni			
Ipertrofia prostatica benigna	Finasteride	2 anni	Placebo	Prevenzione dell'intervento		39 (23 - 111)	
Mal di schiena	Steroidi epidurali			Miglioramento del 75% a 2 mesi		6	
Morsicature di cane	Antibiotici	Breve	Placebo	infezione	0.6 (0.4 - 0.8)	16 (9 - 92)*	13
Neuropatia diabetica	Anticonvulsivanti			Miglioramento del 50%		2.5	
Neuropatie dolorose	Antidepressivi			Miglioramento al 50%		2.5	
Nevralgia del trigemino	Anticonvulsivanti			Miglioramento del 50%		2.5	
Onicomicosi	Terbinafina	12-24 settimane	Griseofulvina	Guarigione a 48 settimane	4.5 (2.3 - 8.8)	2.7 (1.9 - 4.5)*	4
Otite media acuta	Antibiotici	Breve	Nessuna terapia antibiotica né timpanocentesi	Assenza di segni e sintomi a 7-14 giorni	2.9 (1.8 - 4.1)	7	7
Parto prematuro	Corticosteroidi	Prima del parto	Nessun trattamento	Rischio di RDS fetale		11 (8 - 16)*	12
Pediculosi del capo	Permetrina	14 giorni	Placebo	guarigione		1.1 (1.0 - 1.2)*	1
Prevenzione dell'ulcera gastrica da FANS	Misoprostolo	4 settimane	Placebo	Presenza di ulcera gastrica		13	
Prevenzione della frattura di femore	Calcio e vit. D	3 anni	Placebo	Prevenzione di una frattura		20 (13-57)	27
Prevenzione primaria	Vari	5 anni	Nessuna terapia	Prevenzione di un IMA o decesso cerebrovascolare		69 (54 - 99)	28
Prevenzione primaria	Statine	Oltre 4 anni	Placebo	Ogni evento negativo		35 (24 - 63)	
Prevenzione secondaria	Vari	5 anni	Nessuna terapia	Prevenzione di un IMA o decesso cerebrovascolare		16 (13 - 19)	28
Prevenzione secondaria	Statine	2.9 anni in media	Placebo	Ogni evento negativo		11 (10 - 13)	
Sanguinamento di varici esofagee	Legatura endoscopica	Intervento	Scleroterapia	Prevenzione di recidiva		4	6
Stroke: prevenzione primaria MRC 17354 paz, 36-64 anni, PA diastolica 90-109 mmHg	Benzoflurazide Propranololo	5.5 anni	Placebo	Prevenzione di uno stroke ad 1 anno		850	18
Stroke: prevenzione primaria SHEP 4736 paz, 60 anni o più, PA sistolica 160-219 mmHg e diastolica <90 mmHg	Clortalidone Atenololo	4.5 anni	Placebo	Prevenzione di uno stroke ad 1 anno		43 (27 - 95)	19
Stroke: prevenzione primaria STOP 1627 paz, 70-84 anni, PA sistolica 180-230 mmHg e diastolica 90 mmHg opp. diastolica 105-120 mmHg	Atenololo Idroclorotiazide e amiloride Metoprololo Pindololo	4 anni	Placebo	Prevenzione di uno stroke ad 1 anno		34 (20 - 123)	20
Stroke: prevenzione primaria MRC 4396 paz, 65-74 anni, PA sistolica 160-209 mmHg e diastolica <115 mmHg	diuretico e beta-bloccante	5.8 anni	Placebo			70 (36 - 997)	21
Stroke: prevenzione primaria WOSCOPS 6595 maschi, 45-64 anni, colesterolo oltre 6.5mmol/l	Pravastatina	4.9 anni	Placebo			641 (135 - nessun beneficio)	22
Stroke: prevenzione secondaria CATS 1072 paz, 1 settimana - 4 mesi dopo lo stroke	Ticlopidina	2 anni	Placebo			15 (9 - 41)	23
Stroke: prevenzione secondaria SALT 1360 paz, 50-79 anni, TIA o stroke minore nei 3 mesi precedenti	ASA a basse dosi	2.7 anni	Placebo			38 (16 - 85)	24
Stroke: prevenzione secondaria 4S 4444 paz, 35 - 70 anni, angina o IMA, colesterolo 5.5-8.0 mmol/l	Simvastatina	5.4 anni	Placebo			65 (38 - 224)	25
Stroke: prevenzione secondaria ESPS2 6602 paz, 18 anni, TIA o stroke nei precedenti 3 mesi	ASA Dipiridamolo ASA e Dipiridamolo	2 anni	Placebo			37 (20 - 319) 42 (21 - 1000) 8 (13 - 29)	26
Trombosi venosa profonda	Bendaggio compressivo graduato	Non stabilito	Non uso di calze	Episodi di tromboembolia venosa	0.3 (0.2 - 0.4)	9 (7 - 13)*	11
Ulcera peptica	Tripla terapia	6-10 settimane	H2-antagonisti	Eradicazione HP	44 (34 - 56)	1.1 (1.08 - 1.15)	2
Ulcera peptica	Tripla terapia	6-10 settimane	H2-antagonisti	Guarigione a 1 anno	9.4 (6.3 - 14.0)	1.8 (1.6 - 2.1)	2
Ulcera peptica	Tripla terapia	6-10 settimane	H2-antagonisti	Guarigione a 6-10 settimane	5.0 (3.3 - 7.7)	4.9 (4.0 - 6.4)	2
Vomito postoperatorio	Droperidolo	Dose unica	Placebo	Prevenzione a 48 ore in bambini operati per strabismo	2.5 (1.7 - 3.6)	4.4 (3.1 - 7.1)	10

Riferimenti

1. Vander Stichele RH, Dezeure EM, Bogaert MG. Systematic review of clinical efficacy of topical treatments for head lice. *British Medical Journal* 1995; 311:604-8
2. Moore RA. Helicobacter pylori and peptic ulcer. A systematic review of effectiveness and an overview of the economic benefits of implementing that which is known to be effective (<http://www.jr2.ox.ac.uk/Bandolier/bandopubs/hpyl/hp0.html>)
3. Tfelt-Hansen P. Sumatriptan for the treatment of migraine attacks-a review of controlled clinical studies. *Cephalalgia* 1993; 13:238-44
4. Haneke E, Tausch I, Bräutigam M, Weidinger G, Welzel D. Short-duration treatment of fingernail dermatophytosis: A randomized, double-blind study with terbinafine and griseofulvin. *Journal of the American Academy of Dermatology* 1995; 32:72-77
5. Moore A, Collins S, Carroll D, McQuay H. Paracetamol with and without codeine in acute pain: a quantitative systematic review. *Pain* 1997; 70:193-201
6. Laine L, Cook D. Endoscopic ligation compared with sclerotherapy for treatment of esophageal variceal bleeding. A meta-analysis. *Annals of Internal Medicine* 1995; 123:280-7
7. Rosenfeld RM, Vertrees JE, Carr J, Cipolle RJ, Uden DL, Giebink GS et al. Clinical efficacy of antimicrobial drugs for acute otitis media: metaanalysis of 5400 children from thirty-three randomized controlled trials. *Journal of Pediatrics* 1994; 124:355-67
8. Leher P, Comte S, Gamand S, Brown TM. Naftidrofuryl in intermittent claudication: a retrospective analysis. *Journal of Cardiovascular Pharmacology* 1994; 23 (Suppl3):S48-S52
9. Hazell P, O'Connell D, Heathcote D, Robertson J, Henry D. Efficacy of tricyclic drugs in treating child and adolescent depression: a meta-analysis. *British Medical Journal* 1995; 310:897-901
10. Tramer M, Moore A, McQuay H. Prevention of vomiting after paediatric strabismus surgery: a systematic review using the numbers-needed-to-treat method. *British Journal of Anaesthesia* 1995; 75:556-61
11. Wells PS, Lensing AWA, Hirsh J. Graduated compression stockings in the prevention of postoperative venous thromboembolism. *Archives Internal Medicine* 1994; 154:67-72
12. Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994. *American Journal of Obstetrics & Gynecology* 1995; 173:322-35
13. Cummings P. Antibiotics to prevent infection in patients with dog bite wounds: a meta-analysis of randomized trials. *Annals of Emergency Medicine* 1994; 23:535-40
14. Mulrow CD, Cornell JA, Herrera CR, Kadri A, Farnett L, Aguilar C. Hypertension in the elderly. *Journal of the American Medical Association* 1994; 272:1932-8
15. Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. *Lancet* 1988; 2:349-60
16. Silverstein FE, Graham DY, Senior JR, Davies HW, Struthers BJ, Bittman RM et al. Misoprostol reduces serious gastrointestinal complications in patients with rheumatoid arthritis receiving nonsteroidal anti-inflammatory drugs. *American College of Physician*
17. Lancaster T, Silagy C, Gray S. Primary care management of acute herpes zoster: systematic review of evidence from randomized controlled trials. *British Journal of General Practice* 1995; 45:39-45
18. MRC trial of treatment of mild hypertension: principal results. *British Medical Journal* 1985 291: 97-104
19. Prevention of stroke by antihypertensive drug treatment in older persons with isolated systolic hypertension. *Journal of the American Medical Association* 1991 265: 3255-64
20. B Dahlöf et al. Morbidity and mortality in the Swedish trial in old patients with hypertension (STOP-hypertension). *Lancet* 1991 338:1281-5
21. MRC trial of treatment of hypertension in older adults: principal results. *British Medical Journal* 1992 304:405-12
22. J Shepherd et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolaemia. *New England Journal of Medicine* 1995 333:1301-7
23. M Gent et al. The Canadian American ticlopidine study (CATS) in thromboembolic stroke. *Lancet* 1989 i: 1215-20
24. Swedish aspirin low-dose trial (SALT) of 75 mg aspirin as secondary prophylaxis after cerebrovascular ischaemic events. *Lancet* 1991 338:1345-9
25. Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian simvastatin survival study (4S). *Lancet* 1994 344: 1383-9
26. HC Diener et al. European stroke prevention study 2. Dipyridamole and acetylsalicylic acid in the secondary prevention of stroke. *Journal of Neurological Sciences* 1996 143: 1-13
27. MC Chapuy, ME Arlot, F Debouef etc. Vitamin D3 and calcium to prevent hip fractures in elderly women. *New England Journal of Medicine* 1992 327: 1637-42
28. CM Rembold. Number-needed-to-treat analysis of the prevention of myocardial infarction and death by antidyslipidemic therapy. *Journal of Family Practice* 1996 42: 577-86
29. P Madge, J McColl, J Paton. Impact of a nurse-led home management training programme in children admitted to hospital with acute asthma: a randomised controlled study. *Thorax* 1997 52: 223-8

Test di screening

Uno screening viene generalmente messo in atto per uno dei seguenti scopi:

- identificare una patologia in atto, quando questa si trova ancora in fase iniziale ed un intervento precoce offre la possibilità di guarigione o quanto meno di miglioramento della prognosi
- effettuare una ricerca, per valutare una prevalenza e/o il risultato di un intervento preventivo

In base alla popolazione cui si rivolge, può essere:

- selettivo, se indirizzato a un soggetto o a una popolazione a rischio (PAP test, mammografia)
- di massa, se effettuato teoricamente sulla popolazione globale (test di Guthrie per la fenilchetonuria)

Uno screening non ha mai finalità diagnostiche: gli individui risultati positivi devono confermare la diagnosi definitiva prima di essere sottoposti agli opportuni interventi terapeutici o chirurgici del caso.

Sensibilità, specificità, valori predittivi

Uno dei problemi chiave nell'applicazione pratica di uno screening riguarda la capacità dei test utilizzati di classificare correttamente le persone sottoposte. Un test ideale dovrebbe dimostrare la presenza o l'assenza di malattia in ogni individuo, ovvero dovrebbe dare sempre risultati positivi in presenza di malattia e sempre

negativi in sua assenza. Sfortunatamente non esistono test ideali, per cui occorre prendere in considerazione alcuni indicatori che esprimono in termini di probabilità la loro capacità discriminativa: sensibilità, specificità, valore predittivo positivo, valore predittivo negativo ed efficienza.

La **sensibilità** (percentuale di test positivi nei soggetti ammalati, quindi capacità di identificare i veri ammalati) e la **specificità** (percentuale di test negativi nei i soggetti sani, quindi capacità di identificare i soggetti sani) forniscono indicazioni sulla capacità del test di identificare i malati e i sani.

Il **valore predittivo positivo** è calcolato come la percentuale di risultati veri positivi rispetto a tutti i positivi, sia reali che falsi, mentre il **valore predittivo negativo** rapporta i veri negativi a tutti i risultati negativi; le indicazioni fornite riguardano nel primo caso la probabilità che un soggetto con test positivo sia realmente malato e nel secondo la probabilità, con un test negativo, di identificare il soggetto non malato.

L'**efficienza** fornisce la probabilità di ottenere un risultato corrispondente alla realtà, in quanto mette in rapporto gli esami corretti (veri positivi e veri negativi) con il totale di esami eseguiti.

Il test ideale dovrebbe avere valori di sensibilità e specificità entrambi del 100%, ma la condizione è di fatto impossibile, dal momento che esiste, di regola, una parziale sovrapposizione della variabile in esame fra popolazione sana e popolazione malata (o a rischio).

SPECTRUM BIAS

È il fenomeno per il quale sensibilità e specificità di un test variano nelle differenti popolazioni testate a causa di differenze legate al sesso, all'età e ai diversi gradi della malattia.

Un esempio: il test delle urine testato su tutta la popolazione ha una sensibilità di 0.83 (95% CI 0.73-0.91) e una specificità di 0.71 (95% CI 0.66-0.77).

Quando la probabilità a priori di infezione urinaria è alta, la sensibilità del test è alta - 0.92 (0.82-0.98)

Quando la probabilità a priori è bassa, la sensibilità è bassa - 0.56 (0.31-0.79).

Questi esempi si possono applicare a tutti i test.

Per semplificare il problema si può usare un analogo number-needed-to-diagnose (NND).

Per analogia con NNT possiamo definire il NND come il reciproco della frazione di test positivi nel gruppo dei malati meno la frazione dei test positivi nel gruppo senza malattia.

La frazione dei test positivi nel gruppo della malattia è la sensibilità $[VP/(VP+VN)]$ (vero positivo/vero positivo + falso positivo).

La specificità è definita come la proporzione di persone senza malattia che hanno un test negativo. La frazione dei test positivi nel gruppo senza malattia lo possiamo definire come $1 - \text{specificità}$.

$$NND = 1 / [\text{Sensibilità} - (1 - \text{Specificità})]$$

I test per l'*Helicobacter Pylori* (sierologico e breath test) hanno una sensibilità e una specificità del 95%.

$$NND = 1 / [0.95 - (1 - 0.95)] = 1.1$$

Esempio di NND per test diagnostici con spectrum bias:

Il CEA per cancro del colon in pazienti con malattia ha un NND di 4.4 in cancri scoperti precocemente, ma in cancri avanzati diventa 1.6. Tipico esempio di spectrum bias.

Cosa significa tutto questo per il medico pratico?

Davanti a diversi test il medico pratico dovrà familiarizzare con la sensibilità e specificità e scegliere i test con NND più bassi.

Test	Sottogruppo	Sensibilità	Specificità	NND
Es. urina su striscia per infezione	Totale	0.83	0.71	1.8
	Alta probabilità a priori	0.92	0.42	2.9
	Bassa probabilità a priori	0.56	0.78	2.9
Sierologia per infezione da HP	Tutti i pazienti	0.95	0.95	1.1
CEA per screening di cancro colon-retto	Stadio Duke A or B	0.36	0.87	4.3
	Stadio Duke C or D	0.74	0.83	1.8
ECG da sforzo per coronaropatia	Uomini	0.73	0.83	1.9
	Donne	0.57	0.86	2.3
	Sotto 40 anni	0.56	0.84	2.5
	60 anni	0.84	0.70	1.9
Test biochimici per fumatori	CO espirata	0.98	0.92	1.12
	Tiocianato sierico	0.82	0.91	1.37
	Metaboliti urinari della nicotina	0.98	0.94	1.09

Anche in casi limite un test non è in grado raggiungere il 100% di sensibilità e il 100% di specificità.

Un elettroencefalogramma (EEG) piatto è sensibile al 100% in caso di morte in quanto tutti i morti hanno un EEG piatto (sensibilità 100% = TUTTI i test sono positivi se eseguiti a soggetti ammalati). Viceversa, non tutti coloro che improvvisamente presentano un EEG piatto sono morti "definitivamente", quindi l'EEG non è specifico al 100% (specificità 100% = TUTTI i test sono negativi se eseguiti a soggetti sani). La diagnosi di morte in base al "rigor mortis" e al "livor mortis" raggiunge una specificità del 100% (nessun vivo presenta tali segni), ma non si raggiunge la sensibilità del 100% (non TUTTI i cadaveri presentano i due segni, perché potrebbero non averli ancora sviluppati).

Quadro riassuntivo dei criteri di valutazione di un test in relazione a una tabella a due entrate

	Malati	Non malati	Totale
Test positivo	Veri positivi (VP)	Falsi positivi (FP)	VP + FP
Test negativo	Falsi negativi (FN)	Veri negativi (VN)	FN + VN
Totale	VP + FN	FP + VN	VP + FP + VN + FN

Sensibilità = $VP / VP + FN$

Specificità = $VN / FP + VN$

Valore Predittivo (+) = $VP / VP + FP$

Valore Predittivo (-) = $VN / FN + VN$

Efficienza o Accuratezza = $VP + VN / VP + FP + FN + VN$

Se non è possibile avere contemporaneamente massima sensibilità e massima specificità, quale delle due conviene privilegiare? Nella scelta conviene attribuire maggior importanza al valore predittivo positivo o a quello negativo?

La scelta del test deve essere fatta in base alle caratteristiche epidemiologiche della popolazione studiata e alla valutazione del numero di falsi negativi e di falsi positivi che può essere tollerabile: più bassa è la sensibilità più alta sarà la percentuale di falsi negativi (negativi al test, ma affetti dalla patologia indagata), più bassa è la specificità più alta sarà la quota di falsi positivi (positivi al test, ma privi della stessa).

Inoltre è necessario considerare la prevalenza della patologia, la sua gravità, il costo del test e i vantaggi nonché le probabilità di successo di un trattamento precoce.

Una più alta sensibilità (preferibilmente del 100%) è richiesta nel caso di patologia grave che deve necessariamente essere diagnosticata in quanto curabile, se nel contempo i risultati falsi positivi noti

comportano danni psicologici o economici per il paziente. In un'indagine generalizzata, se la malattia in esame ha una bassa prevalenza e la sensibilità del test è del 100%, si raggiunge lo scopo di identificare tutti i casi di malattia. Per contro, si avrà un elevato numero di diagnosi falso positive, inconveniente ovviabile, a seconda dei casi, con la ripetizione del medesimo test o di un altro più specifico.

Una più alta specificità (preferibilmente del 100% è richiesta nel caso di malattie gravi e ad alta letalità: è meno dannoso non identificare una sclerosi multipla, che sarà comunque identificata in seguito, piuttosto che prospettare una tale patologia a una persona sana con possibili danni psicologici o economici.

Il valore predittivo di un test positivo è notevolmente influenzato dalla prevalenza della malattia, che può risultare più importante della sensibilità e della specificità del test. Il valore predittivo positivo, a pari sensibilità e specificità, risulta massimo per una prevalenza del 50% e diminuisce in maniera esponenziale con il decrescere della prevalenza. Anche per una predittività negativa entra in gioco la prevalenza della situazione in esame. Un alto valore predittivo positivo è essenziale quando un risultato falso positivo potrebbe avere gravi conseguenze se per una data malattia sono noti solo trattamenti pericolosi o invalidanti: una loro applicazione a un individuo non malato porterebbe a conseguenze disastrose.

D'altronde il limite di separazione fra valore normale e non normale viene stabilito arbitrariamente da considerazioni cliniche derivanti dalla conoscenza della malattia.

Quando vengono usati test quantitativi è possibile variare la positività e la negatività cambiando il livello di soglia. Per l'anemia delle donne può essere accettabile un valore soglia sia di 11, sia di 12 grammi di emoglobina per 100 ml di sangue intero. Un limite di 12 grammi potrebbe assicurare una maggior sensibilità ma una minor specificità, poiché una più larga proporzione di soggetti normali potrebbe essere classificata come anemica.

D'altra parte se il limite è di 11 grammi la sensibilità diminuisce, mentre la specificità aumenta. Questo limite di 11 grammi potrebbe non evidenziare alcuni individui anemici con valori di emoglobina ai limiti, ma eviterebbe la classificazione errata di soggetti sani.

Una elevata efficienza è richiesta invece nel caso di una malattia grave, ma curabile e comunque nel caso in cui risultati falsi positivi e falsi negativi portino a situazioni pericolose. L'infarto del miocardio e il diabete mellito possono essere trattati, ma risultano pericolose sia diagnosi falso positive, sia falso negative.

Spesso risulta più utile considerare il valore predittivo dei risultati ottenuti da più test, mediante una loro combinazione in parallelo (per cui vengono considerati positivi i soggetti che presentano almeno un test positivo e negativi i soggetti per i quali tutti i test sono negativi), o in serie, e in questo caso sarà considerato positivo un soggetto che risulta positivo a tutti i singoli test.

La scelta del criterio e della sequenza migliore dipendono dalla situazione, dalla sensibilità e dalla specificità dei singoli test e dalla loro combinazione. In una stessa situazione i test eseguiti in parallelo permettono di rilevare un numero più elevato di ammalati (maggiore sensibilità), ma portano anche a un notevole incremento di falsi positivi (perdita in valore predittivo) e alla esigenza di sottoporre tutti i pazienti a tutti i test (costi umani ed economici).

Al contrario, test eseguiti in serie risultano meno sensibili, ma acquistano in specificità e risultano più economici in quanto vengono sottoposti a un test solo i soggetti risultati positivi al test precedente (naturalmente deve essere stabilita la sequenza ottimale usando un modello predittivo). L'applicazione delle percentuali di sensibilità e di specificità dei test in serie è valida solo nel caso in cui i test siano indipendenti tra loro; nella situazione contraria il valore predittivo calcolato risulta più elevato di quello reale, con un errore che dipende dalla correlazione dei test.

GLI INDICI DI TENDENZA CENTRALE

Media aritmetica semplice

La media aritmetica semplice è l'indice di tendenza centrale più utilizzato.

La media si ottiene mediante la sommatoria di tutte le osservazioni (x) diviso il loro numero (N):

$$m = \bar{x} = \frac{\sum x_i}{N}$$

Al numeratore il segno di sommatoria (sigma greco maiuscolo) significa "somma di".

Limitazioni della media aritmetica

- dati non quantitativi
- differenti ordini di grandezza
0.7 10 20 37.9 400
- valori estremi indefiniti
<8 5 9 10 >50
- distribuzioni di frequenza con classi aperte

Media aritmetica ponderata

In presenza di misure uguali è prevista la ripetizione della sommatoria per un numero di volte pari alla frequenza con cui compare.

In questo caso la media può essere calcolata moltiplicando le singole misure per la propria frequenza f .

$$\bar{x} = \frac{\sum f_i * x_i}{\sum f_i} = \frac{\sum f_i * x_i}{N}$$

Il rapporto f_i/N è la frequenza relativa di ciascuna modalità assunta dalla variabile e rappresenta il peso da attribuire a ciascun valore.

La media aritmetica della mortalità in Italia non si può intendere come media semplice dei dati regionali. Con una distribuzione in classi è previsto l'uso della media ponderata.

In questo modo verrebbero messe sullo stesso piano regioni con diverso numero di abitanti. Va quindi applicata la media ponderata.

Esempio di uso della media ponderata (tassi grezzi di mortalità, 1991)

Regione	morti x 1000 ab.	Popolazione
Piemonte	11.5	4.290.412
Valle d'Aosta	10.7	115.397
Lombardia	9.3	8.831.264
Trentino Alto Adige	9.3	886.914
Veneto	9.5	4.363.157
Friuli Venezia Giulia	12.6	1.193.520
Liguria	13.8	1.668.078
Emilia Romagna	11.6	3.899.170
Toscana	11.7	3.510.114
Umbria	11.1	804.054
Marche	10.2	1.427.666
Lazio	9.0	5.031.230
Abruzzo	9.9	1.243.690
Molise	9.9	327.893
Campania	8.1	5.589.587
Puglia	7.5	3.986.430
Basilicata	8.4	605.940
Calabria	7.8	2037686
Sicilia	8.9	4.961.383
Sardegna	8.1	1.637.705
Media semplice	9.95	
Media ponderata	9.66 -> 9.7	
Italia	9.7	56.411.290

Proprietà della media aritmetica

1. La somma algebrica degli scarti della media è uguale a zero:

$\sum(x - \bar{x}) = 0$ Essendo un valore centrale gli scarti positivi vengono bilanciati da quelli negativi.

2. L'espressione $\sum(x-A)^2$ dove A è un valore arbitrario scelto a piacere assume il minimo valore per $A=\bar{x}$

3. Date più medie e le singole frequenze con cui sono state calcolate, la media generale può essere calcolata come media ponderata delle medie: $\bar{x} = \frac{\sum f_i * \bar{x}_i}{\sum f_i}$

4. Per semplificare i calcoli, si può sottrarre da ciascun dato un valore arbitrario A (media arbitraria) ottenendo una nuova variabile $b = x - A$. La media risulta:

$$\bar{x} = A + \frac{\sum(x-A)}{N} = A + \bar{b}$$

e per dati ripetuti o in classi diventa:

$$\bar{x} = A + \frac{\sum f * (x-A)}{\sum f} = A + \bar{b}$$

Esempio: la media aritmetica dei valori 102 107 104 102 104 98 risulta

$$\bar{x} = \frac{\sum x}{N} = \frac{102+107+104+102+104+98}{6} = 102,8$$

poiché ci sono valori ripetuti si può applicare la media ponderata

$$\bar{x} = \frac{\sum f_i * x_i}{\sum f_i} = \frac{2*102+107+2*104+98}{2+1+2+1} = 102,8$$

per un calcolo manuale, con dati con lo stesso ordine di grandezza, le operazioni possono essere semplificate utilizzando la media arbitraria: posto $A=100$, le differenze $b_i = (x_i - A)$ diventano 2 7 4 2 4 -2. Se alla media di questi valori si aggiunge il valore arbitrario $A=100$ si ottiene la media:

$$\bar{b} = \frac{2*2+7+2*4-2}{6} = \frac{17}{6} = 2,8$$

$$\bar{x} = A + \bar{b} = 100 + 2,8 = 102,8$$

Media geometrica

In certi casi quando i dati sono distribuiti su diversi ordini di grandezza è bene utilizzare la media geometrica.

La media aritmetica di 10 100 e 1000 è 370 troppo spostata verso il valore più alto e quindi non da un valore reale di centralità.

La media geometrica è data da:

$$MG = \sqrt[N]{x_1 * x_2 * \dots * x_N} = \sqrt[N]{\prod_{i=1}^N x_i}$$

Cioè la radice ennesima del prodotto delle N osservazioni.

In caso di osservazioni ripetute, o in classi, la formula viene adattata come media ponderata.

$$MG = \sqrt[\sum f_i]{x_1^{f_1} * x_2^{f_2} * \dots * x_i^{f_i}} = \sqrt[\sum f_i]{\prod x_i^{f_i}}$$

Media armonica

E' di uso meno frequente ed è definita come reciproco della media aritmetica del reciproco delle misure:

$$MA = \frac{1}{\sum \frac{1}{x_i}} = \frac{N}{\sum \frac{1}{x_i}}$$

La media armonica permette di tenere conto anche di valori infiniti (il reciproco di ∞ è 0) che impedirebbero il calcolo di altre medie analitiche.

Trova applicazione nella valutazione di tempi di risposta in prove di tossicità, sopravvivenza post-operatoria e in genere quando un fenomeno si dovrebbe esaurire in un arco di tempo definito.

Moda

Nel caso di misure nominali l'unico criterio per sintetizzare la tendenza centrale consiste nell'individuare il dato che compare maggiormente e che viene definito moda.

Essa restituisce il valore più ricorrente, o ripetitivo, di una matrice o di un intervallo di dati. Analogamente alla mediana, la moda è una misura relativa alla posizione dei valori.

La formula è la seguente:

$$\text{Moda} = L_1 + \frac{\Delta_1}{(\Delta_1 + \Delta_2)} * c$$

dove

L_1 e c sono il confine inferiore e l'ampiezza della classe modale.

Δ_1 e Δ_2 sono le differenze rispettivamente tra la frequenza della classe modale e la precedente (Δ_1) e tra la frequenza della classe modale e la successiva (Δ_2)

c è l'ampiezza di classe.

1 1 2 2 3 3 4 4 4 5 5 6 6 6 6 8 10 10 40 40 40

Sopravvivenza di 21 pazienti dopo la diagnosi di neoplasia polmonare (mesi).

moda = 6, unica misurazione che compare 4 volte.

Mediana

Nel caso di misure ordinali è possibile utilizzare la mediana. Tale indice separa in due parti numericamente uguali le osservazioni. Se si hanno N osservazioni la mediana corrisponde al valore in posizione $(N+1)/2$. Se il numero di osservazioni è dispari la mediana coincide con l'osservazione di mezzo; se N è pari non esiste una osservazione centrale e la mediana sarà data dalla media tra le due osservazioni centrali $N/2$ e $(N+1)/2$.

Indici di tendenza centrale utilizzabili con le varie scale di misura

Scala di misura

indici

Nominale

moda

Ordinale

moda, mediana

Intervallare

moda, mediana, media

La mediana presenta diversi svantaggi rispetto alla media:

1. non tiene conto esattamente della grandezza delle osservazioni e quindi sciupa delle informazioni

2. si presta meno facilmente ad una elaborazione matematica e perciò è meno utilizzabile per valutazioni statistiche elaborate.

La mediana è tuttavia utile nell'indagine statistica descrittiva riferita a misure ordinali e può essere validamente utilizzata anche per dati quantitativi quando l'utilizzo della media sia sconsigliato. La mediana, non risentendo dei valori estremi della serie, è utilizzabile in presenza di valori della variabile molto diversi e persino di valori estremi indeterminati o infiniti.

Un'estensione del concetto di mediana è costituito dai quantili, un termine generico che individua una serie di indicatori che hanno in comune la prerogativa di suddividere in modo definito una serie ordinata. Si utilizzano i quartili (Q_i) i decili (D_i) ed i percentili (P_i) che suddividono rispettivamente in 4, 10 e 100 le misure della variabile. Così il primo quartile separerà il 25% delle osservazioni da tutte le altre ed il terzo il 75% delle medesime osservazioni.

E' da notare che il secondo quartile, il quinto decile ed il cinquantesimo percentile comprendono la mediana.

Per una serie di dati i quantili si calcolano nel seguente modo:

$$Q_i = \frac{i}{4}(N+1)$$

$$D_i = \frac{i}{10}(N+1)$$

$$P_i = \frac{i}{100}(N+1)$$

Esempio di calcolo di moda, mediana e di alcuni quantili in una distribuzione in classi

Classe (cm)	f_a	f_{cum}
150-154	2	2
155-159	6	8
160-164	11	19
165-169	18	37
170-174	25	62
175-179	13	75
180-184	7	82
Σ	82	

Moda:

classe modale = 170 -174

$$\text{valore} = L_1 + \frac{\Delta_1}{(\Delta_1 + \Delta_2)} * c = 169.5 + \frac{25-18}{(25-18)+(25-13)} * 5 = 171.3$$

Mediana:

posizione = $N/2 = 82/4 = 41$ (classe 170 -174)

$$\text{valore} = L_1 + \frac{\frac{N}{2} - f_{cum}}{f_{med}} * c = 169.5 + \frac{41-37}{25} * 5 = 170.3$$

Nono decile (D_9):

posizione = $N*9/10 = 82*9/10 = 73.8$

$$\text{valore} = L_1 + \frac{N * \frac{9}{10} - f_{cum}}{f_{dec}} * c = 174.5 + \frac{73.8-62}{13} * 5 = 179$$

Primo quartile (Q_1):

posizione = $N*1/4=82*1/4 = 20.5$ (classe 165-169)

$$\text{valore} = L_1 + \frac{N * \frac{1}{4} - f_{cum}}{f_{quan}} * c = 164.5 + \frac{20.5 - 19}{18} * 5 = 164.9$$

Quinto percentile (P_5):

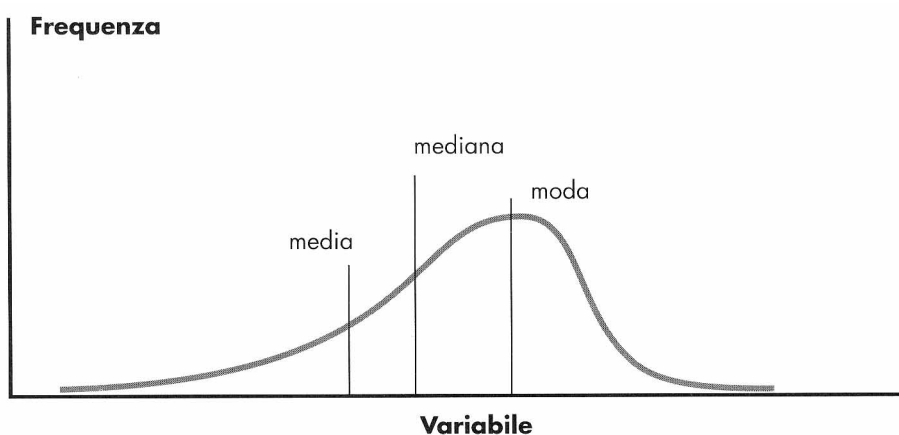
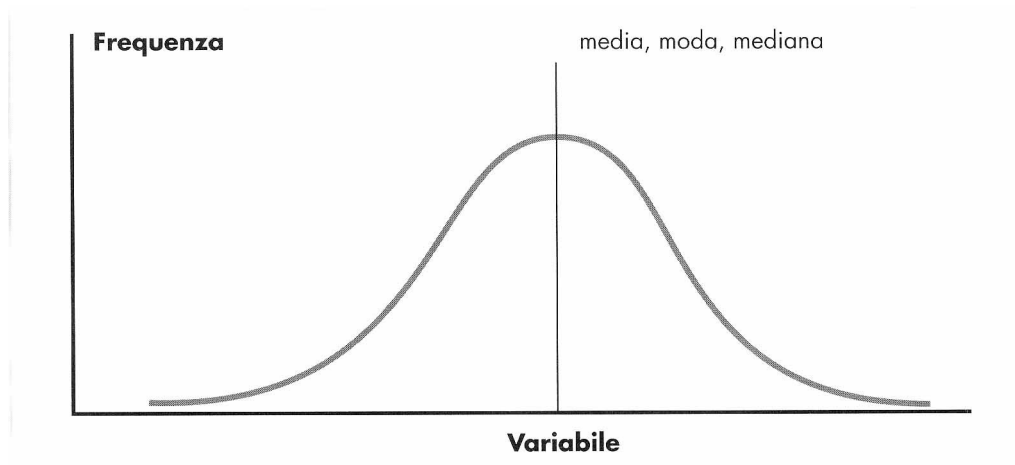
$$\text{posizione} = N * 5 / 100 = 82 * 1 / 20 = 4.1 \text{ (classe 155-159)}$$

$$\text{valore} = L_1 + \frac{N * \frac{5}{100} - f_{cum}}{f_{perc}} * c = 154.5 + \frac{4.1 - 2}{6} * 5 = 156.3$$

Indici di tendenza centrale e distribuzione dei dati

Moda, media e mediana sono gli indici più frequentemente utilizzati. Nell'ambito di misure continue (quantitative) possono essere calcolate tutte e tre.

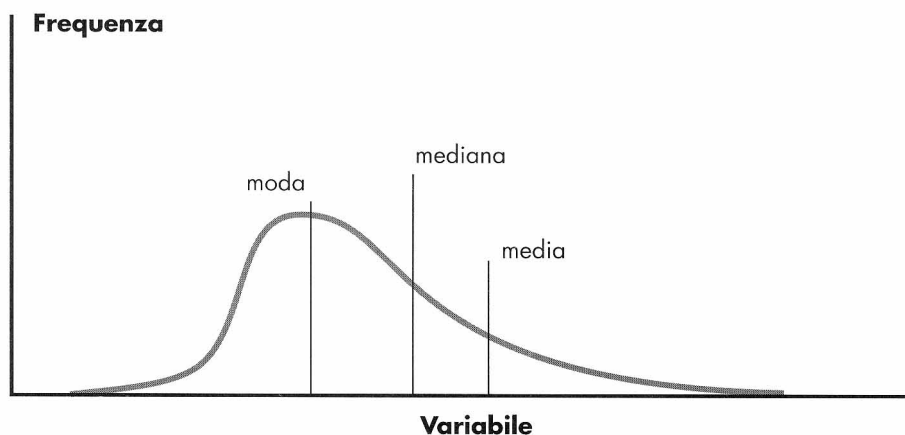
Per distribuzioni simmetriche, unimodali, quali ad esempio la curva normale di Gauss, le tre misure coincidono.



Relazione tra gli indici di tendenza centrale per una distribuzione asimmetrica a sinistra

Nella pratica, una perfetta simmetria è rara, per cui il riscontro di valori simili per i tre indici è considerato sufficiente per ritenere che le osservazioni siano distribuite in modo simmetrico intorno alla media.

Quanto più i tre indici si differenziano tra loro, tanto più la distribuzione diventa asimmetrica. In tal caso gli indici si presentano in ordine alfabetico (media, mediana, moda) quando la coda più prolungata della curva è a sinistra ed in ordine inverso a quello alfabetico (moda, mediana, media) quando è a destra.



Relazione tra gli indici di tendenza centrale per una distribuzione asimmetrica a destra.

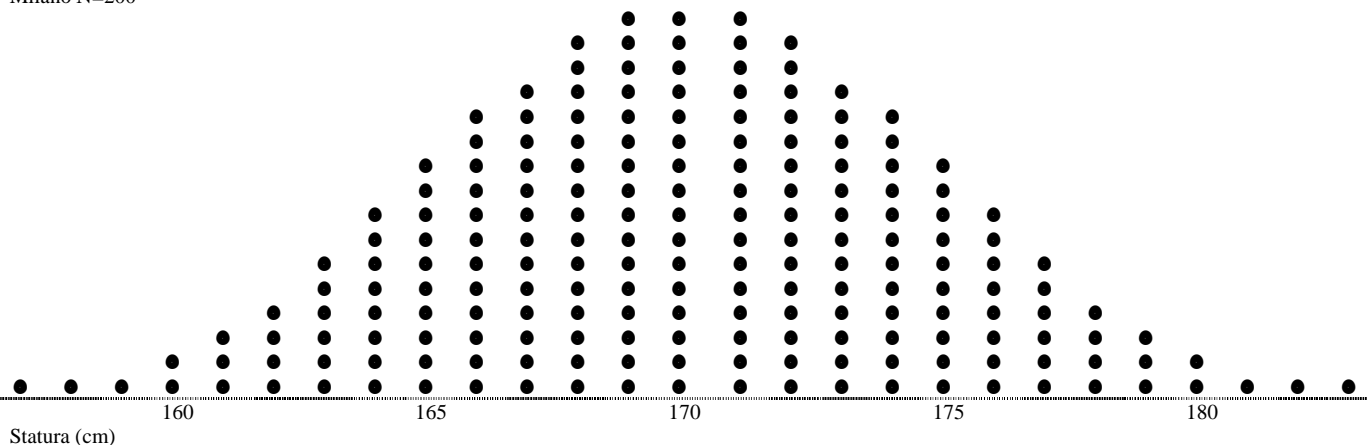
GLI INDICI DI DISPERSIONE

Immaginiamo di valutare l'altezza dei Milanesi e dei Cagliariitani. Prendiamo un campione rappresentativo di 200 persone a Milano e di 150 a Cagliari e misuriamone l'altezza. Riportiamo il risultato su di un grafico, e calcoliamo la media delle altezze mediante la formula che abbiamo già visto:

$$M = \frac{\sum x}{N}$$

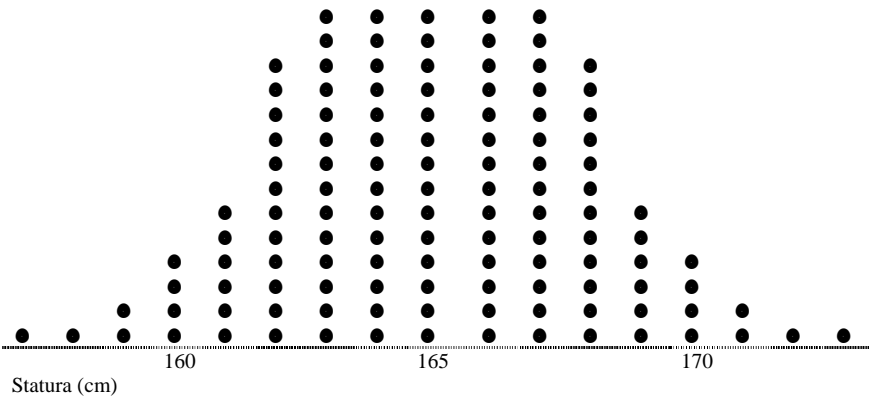
Otterremo 170 cm per Milano e 165 cm per Cagliari.

Milano N=200



Cagliari N=150





Lo studio di un fenomeno biologico, appare più completo se, oltre alla tendenza centrale, si è in grado di misurare anche la capacità di un fenomeno di manifestarsi in modo diverso. Bisogna cioè introdurre il concetto di **indici di dispersione**, da associare alla media, per indicare la variabilità intorno ad essa.

Un primo parametro che possiamo calcolare è la **Varianza (σ^2)** che rappresenta la media dei quadrati degli scostamenti, o scarti, dalla media:

$$\sigma^2 = \frac{\sum(x - \mu)^2}{N}, \text{ dove:}$$

x = valori assunti da ogni elemento della popolazione

μ = media della popolazione

N = numero di elementi della popolazione

In statistica sanitaria è di uso comune usare al denominatore invece di N (N-1) che viene chiamato “gradi di libertà”. Da notare che l'unità di misura della varianza è il quadrato perché la sommatoria degli scarti dalla media sarebbe uguale a zero (vedi le proprietà della media).

Per i Milanesi la varianza è 26.6 e per i Cagliariitani la varianza è 6.9.

Questi dati riassumono che la variabilità dei dati è maggiore nei Milanesi rispetto ai Cagliariitani.

Per esprimere meglio la variabilità dei dati si usa più comunemente la **Deviazione Standard** o DS usando la radice quadra della varianza.

Ossia:

$$DS = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

La DS dei Milanesi è 4.86 e per i Cagliariitani è 2.64.

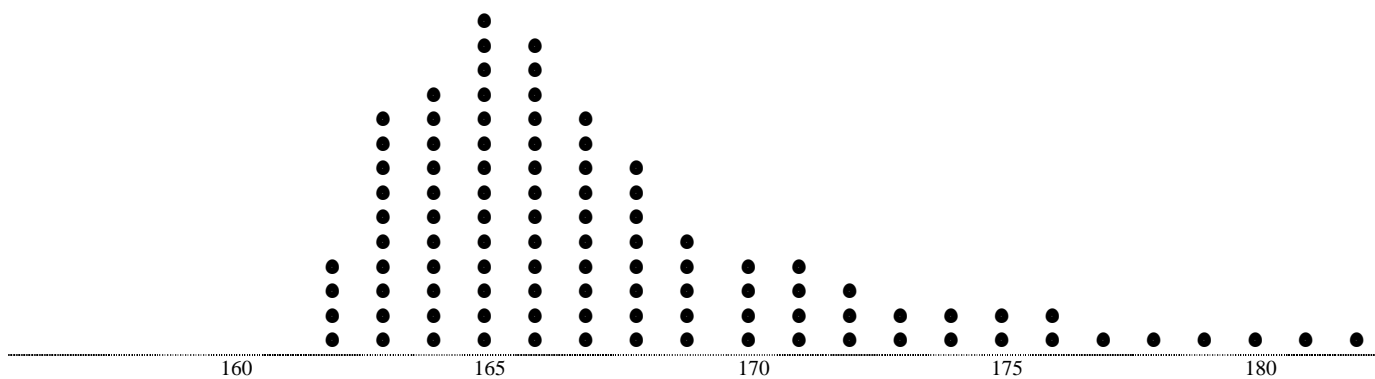
	N° Popolazione	Media	DS
Milanesi	200	170	4.86
Cagliariitani	150	165	2.64

La distribuzione delle stature nelle due città assume una forma analoga, così che il 68% delle stature cade entro una deviazione standard e il 95% entro 2 deviazioni standard dalla media.

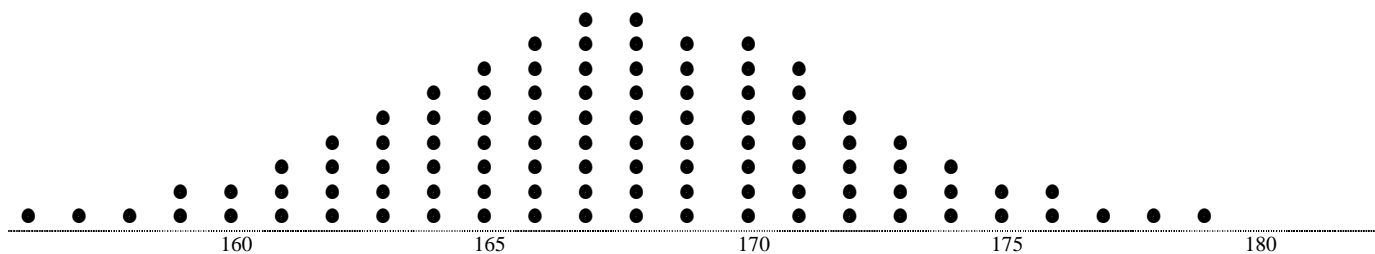
Come si vede i dati nel grafico hanno una forma a campana e questo comportamento si verifica così frequentemente che gli statistici chiamano questa distribuzione **Distribuzione Normale** o **Gaussiana**.

A questo punto immaginiamo di calcolare le altezze degli abitanti di Firenze prendendo un campione di 100 Fiorentini.

N=100 Media=167.6 DS=4.5



N=100 Media=167.6 DS=4.5



Al contrario degli abitanti delle altre città, i Fiorentini mostrano una distribuzione asimmetrica intorno alla media. I pochi individui che sono molto più alti degli altri rendono più elevate media e DS, in modo da indurci a pensare che la maggior parte degli individui sia più alta e la variabilità tra le stature più alta rispetto a quanto si verifica nella realtà. La seconda figura qui sopra mostra un campione di 100 individui la cui distribuzione è normale o Gaussiana con media e DS uguale a quella del campione di Fiorentini. Come si nota c'è una notevole differenza.

È evidente che, quando la distribuzione non è normale, come nella seconda figura, media e DS non riassumono la distribuzione delle stature con la stessa efficacia riscontrata quando queste seguono una distribuzione normale.

Una alternativa a riassumere questi dati è l'uso della mediana. Nel capitolo precedente abbiamo imparato la formula per il calcolo della mediana. Essa esprime un valore rispetto al quale metà dei valori degli individui risulta superiore e metà inferiore.

Fig A

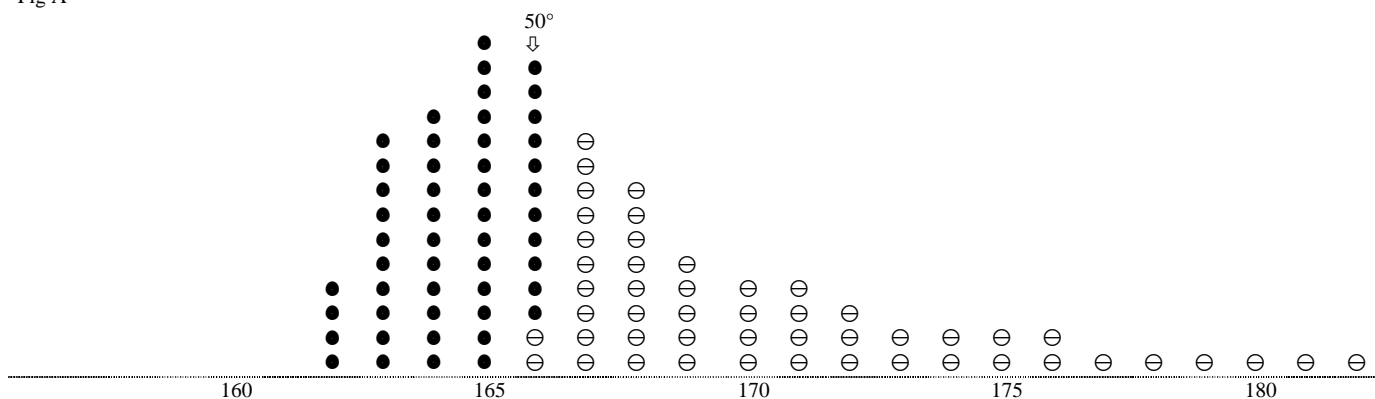
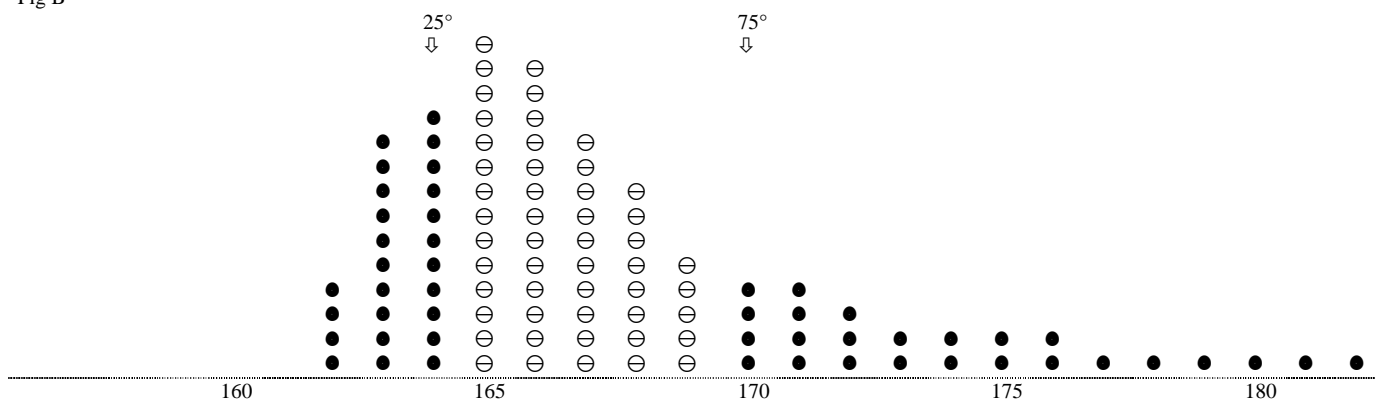


Fig B

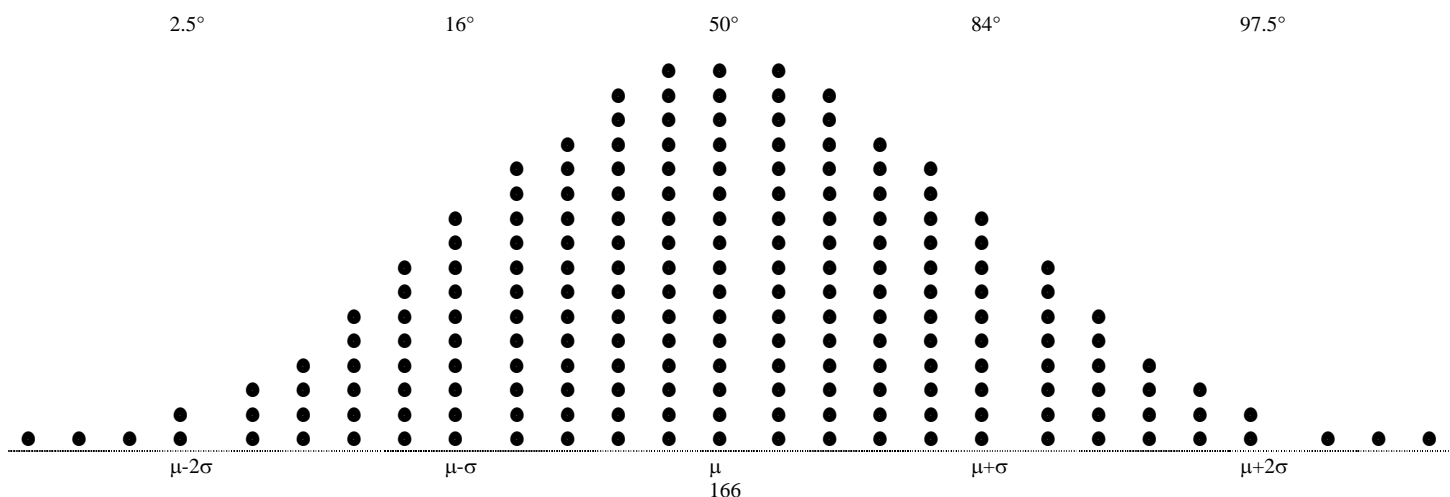


La figura A mostra come la metà dei Fiorentini sia più alta di 165 per cui 165 è la mediana. Dato che il 50% dei valori cade al di sotto della mediana, questa è anche chiamata 50° percentile. Per dare una indicazione sulla dispersione delle stature nella popolazione, si riporta il valore che separa il 25% inferiore e il valore che separa il 25% superiore. Questi valori vengono definiti 25° e 75° percentile.

Dalla figura B si nota che i due valori siano eguali a 164 e 170. Questi sono i percentili più usati, ma si possono indicare anche il 5° e il 95° percentile.

L'uso dei percentili dell'altezza e peso è di uso comune nel bilancio di salute in età pediatrica in quanto valuta la distribuzione del soggetto rispetto alla distribuzione normale e la sua curva di distribuzione nel tempo.

Occorre ricordare che, in una popolazione che segue una distribuzione normale, il 95% dei valori è compreso entro due deviazioni standard dalla media e il 68% entro una deviazione standard.



La figura mostra come la metà dei Fiorentini sia più alta di 166; di conseguenza 166 è la mediana. Dal momento che il 50% dei valori cade al di sotto della mediana, questa è anche chiamata 50° percentile.

Una volta calcolate la media e la DS dobbiamo quantificare l'accuratezza di queste stime. Per quantificare l'accuratezza dobbiamo calcolare i loro **errori standard**.

Per errore standard della media si intende il grado di certezza o di incertezza che la media sia la vera media. Per comprendere meglio il concetto farò un esempio.

Immaginiamo di estrarre dal campione dei 200 Milanesi venti campioni di 10 Milanesi e calcoliamo per ogni campione la media e la DS.

Quando la variabile studiata è la somma di molte variabili, la sua distribuzione tenderà ad essere gaussiana. Quindi ha senso descrivere i dati in termini di media e DS. Dato che la media dei venti valori, riportati in figura, è la media delle venti medie, la indicheremo come $\bar{X}_{\bar{x}}$, mentre la DS delle venti DS la indicheremo come $s_{\bar{x}}$.

Si immagina che dal calcolo esca $\bar{X}_{\bar{x}} = 170$ e $s_{\bar{x}} = 2$

$s_{\bar{x}}$ viene chiamato **errore standard della media**.

Poiché la precisione di stimare la media cresce al crescere delle dimensioni del campione, l'errore standard della media diminuisce al crescere delle dimensioni del campione. Viceversa, quanto maggiore sarà la variabilità della popolazione, tanto maggiore sarà la DS. L'errore standard cresce al crescere della DS.

$$\text{Errore standard} = \text{ES} = \frac{\sigma}{\sqrt{n}}$$

Come si nota, al crescere della DS, cioè della variabilità del campione, aumenta anche l'errore standard e al crescere della dimensione del campione diminuisce la DS e l'errore standard.

La DS e l'errore standard sono due concetti decisamente diversi, ma spesso vengono interpretati o usati come sinonimi. Molte volte i ricercatori riassumono i dati con l'errore standard della media in quanto è inferiore alla DS. Questo comporta un miglior risultato quando in realtà non lo è.

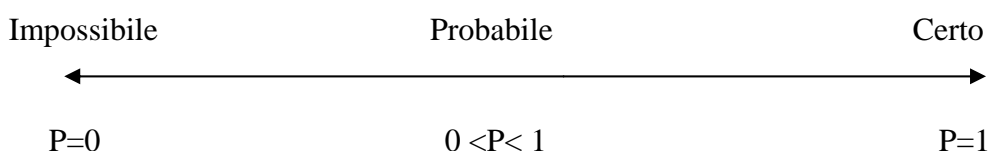
Immaginiamo, per chiarire meglio il concetto, che il volume respiratorio massimo di 10 soggetti sia di 4 litri e la DS sia di 1 litro. Ciò significa che il 95% dei soggetti del campione cade entro due DS dalla media, ovvero è assai improbabile trovare soggetti con volume respiratorio massimo oltre i 6 ml e sotto i 2 ml. Se usiamo, al posto della DS, l'errore standard, che è ad esempio di 0.2 ml, avremo che il 95% dei soggetti del campione ha un volume respiratorio massimo compreso tra 3.8 e 4.2 ml. Ovviamente questo è sbagliato.

TEST DI SIGNIFICATIVITA' E INTERVALLI DI CONFIDENZA

Se indichiamo con P la probabilità che un dato evento si verifichi, l'evento contrario, cioè il fatto che l'evento non si verifichi (Q) è correlato in modo complementare:

$$Q = 1 - P$$

La probabilità dell'evento viene indicata da un numero compreso tra 0 e 1 o, nel caso si utilizzi la frequenza percentuale, tra 0 e 100. I valori compresi in tale intervallo esprimono l'incertezza dell'evento. Un evento impossibile avrà probabilità 0, uno certo probabilità 1 ed uno probabile un valore compreso tra tali estremi.



Immaginiamo di studiare una sequenza casuale di numerose nascite.

M M F M F F M F M M M F M F M M F F M F F F M F

MMFFMFMFMFFMFMFFMFM FFMM

Si può notare come al ripetersi delle osservazioni la frequenza relativa di ciascun evento (M o F) tenda a stabilizzarsi sul valore 0.5 (valori di P o Q)

Osservazione	Maschio	frequenza	p	Q
1	si	1.00	1.00	0
2	si	1.00	1.00	0
3	no	0.66	0.66	0.34
4	si	0.75	0.75	0.25
5	no	0.60	0.60	0.40
6	no	0.50	0.50	0.50
7	si	0.57	0.57	0.43
8	no	0.50	0.50	0.50
9	si	0.55	0.55	0.45
10	si	0.60	0.60	0.40
11	si	0.63	0.63	0.37
<hr/>				
24	no	0.50	0.50	0.50

Immaginiamo che la frequenza dei maschi nella popolazione studiata sia 0.51.

Qual'è la probabilità di avere solo figli maschi oppure due maschi e due femmine etc. ? La probabilità che il primo figlio sia femmina è uguale a $1 - P = Q$, cioè $1 - 0.51 = 0.49$; analogamente la probabilità che i primi due figli siano femmine, per il principio del prodotto, sarà $(0.49) \times (0.49) = (0.49)^2 = 0.24$ il che equivale a dire che 24 famiglie su 100, ossia circa 1 famiglia su 4 ha due figlie femmine.

Una famiglia con due maschi e due femmine può originarsi con sei possibili sequenze casuali.

Ognuna di queste combinazioni ha per il principio sopra ricordato, una probabilità uguale a $(0.49)^2 \times (0.51)^2 = 0.062$ e per il principio della somma la probabilità totale di avere 2 maschi e 2 femmine sarà data da: $0.062 + 0.062 + 0.062 + 0.062 + 0.062 + 0.062 = 0.062 \times 6 = 0.374$

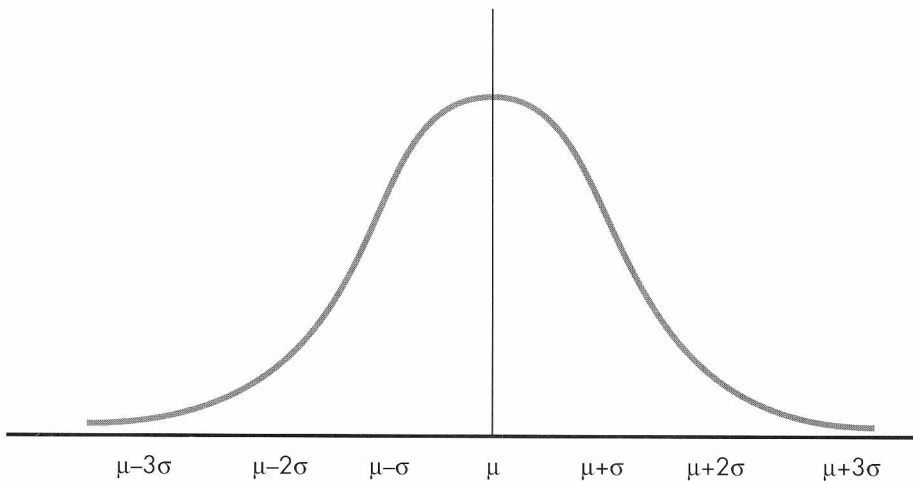
Così facendo otterremo:

M	F	p
o	4	$(0.49)^4 = 0.0576$
1	3	$4 (0.49)^3 (0.51) = 0.2400$
2	2	$6 (0.49)^2 (0.51)^2 = 0.3747$
3	1	$4 (0.49) (0.51)^3 = 0.2600$
4	o	$(0.51)^4 = 0.0677$
		TOTALE = 1.00

Se si rappresentasse in modo grafico la distribuzione si potrebbe osservare che, all'aumentare del numero di osservazioni tende alla simmetria.

La distribuzione più importante in campo statistico è definita distribuzione o curva di Gauss, gaussiana o distribuzione normale.

Rappresentazione grafica di una distribuzione normale - curva di Gauss.



Per poter rappresentare la distribuzione gaussiana occorre possedere due informazioni: il valore della media (μ) e della deviazione standard (σ).

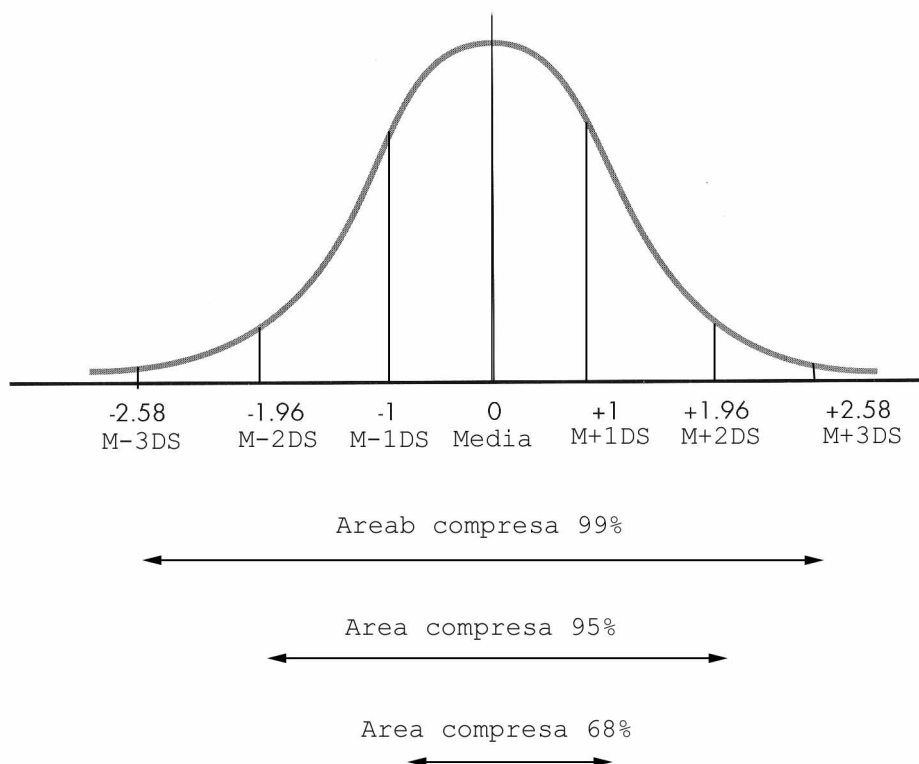
Al variare della media la distribuzione si sposta lungo l'asse delle ascisse ed in base alla deviazione standard si allarga o si restringe. Il valore della deviazione standard assume anche il ruolo di unità di misura dell'asse dell'ascisse ed ogni valore di x (la variabile) differisce da μ in base ad una frazione di σ .

L'area sottesa alla curva di Gauss per valori da $-\infty$ a $+\infty$ è uguale a 1 in quanto la distribuzione esaurisce lo spazio campione della variabile. Esiste perciò la certezza, cioè la probabilità del 100%, che una qualsiasi misura sia inclusa nella distribuzione. Calcolando l'area sottesa ad una porzione di curva si può valutare la probabilità di osservare una data misura. Per facilitare il calcolo è stata introdotta una nuova variabile detta z quale misura dell'asse delle ascisse (distribuzione gaussiana standardizzata):

$$z = \frac{x - \mu}{\sigma}$$

z assume valori di 0 per $x = \mu$ e valori di 1 nei punti di flesso della curva.

Distribuzione gaussiana standardizzata.



Correlazione tra varie misure che identificano punti precisi dello gaussiana.

Gaussiana	Gaussiana Standardizzata	Percentili
M	0	50
$M - \sigma$	-1	16
$M - 2\sigma$	-1.96	2.5
$M - 3\sigma$	-2.58	0.5
$M + \sigma$	+1	84
$M + 2\sigma$	+1.96	97.5
$M + 3\sigma$	+2.58	99.5

Ai fini della nostra trattazione il calcolo di porzioni di area non è rilevante e può comunque essere effettuato con opportune tabelle di distribuzione di z (Tabella 1).

Esistono però alcune posizioni o porzioni della superficie della curva che vale la pena di descrivere più in dettaglio in relazione alla loro notevole importanza ai fini del ragionamento statistico. L'interesse statistico, come vedremo più avanti, si incentra infatti su alcuni valori, arbitrariamente scelti, ma ormai universalmente accettati, che sono ritenuti espressione di un margine di errore accettabile della stima.

Valori di probabilità caratteristici della distribuzione normale.

Deviata	deviata standardizzata	area compresa	area esclusa in entrambe le direzioni	area esclusa in una direzione
$M \pm \sigma$	± 1	0.68	0.32	0.1
$M \pm 2\sigma$	± 1.96	0.95	0.05	0.025
$M \pm 3\sigma$	± 2.58	0.99	0.01	0.005

Poiché l'area compresa tra due misurazioni equivale alla probabilità di osservare un determinato valore della variabile e l'area esclusa alla probabilità di rilevare una maggiore deviazione dalla media, ne consegue che

entro l'intervallo compreso tra il doppio della deviazione standard (2σ) equivalente a $1.96z$ ricadono il 95% delle osservazioni e che solo il restante 5% dei casi sono esclusi ed equamente divisi in maggiori di $M + 2\sigma$ ed inferiori di $M - 2\sigma$. Il 2.5% delle osservazioni costituiscono cioè la porzione di area posta a destra di $M + 2\sigma$ ed il 2.5% la porzione posta a sinistra di $M - 2\sigma$.

Fatte queste premesse possiamo dire che lo scopo della statistica è di rispondere a domande precise come:
Il farmaco A è migliore del farmaco convenzionale o al placebo?
La chemioterapia per un certo tumore aumenta la sopravvivenza rispetto al placebo?
Esiste una correlazione tra un fattore di rischio e una determinata malattia? ecc.

Per rispondere a domande precise gli autori degli studi hanno usato a dismisura il test p di significatività. Spesso si vedono lavori che citano solo il test p per validare le loro ipotesi senza alcun accenno ad altri parametri statistici. Il motivo dell'eccessivo uso del test p consiste nel ritenere di identificare una semplice risposta: sì o no. Ma questo è chiaramente sbagliato e assai limitativo.

Quando si esegue uno studio dove si paragona un trattamento A rispetto ad un trattamento B, alla fine si esegue un test statistico e si dice che il farmaco A è statisticamente più efficace di B con $p < 0.05$. Se $p > 0.05$ si dice che il farmaco A non ha raggiunto la significatività statistica per dichiararlo più efficace di B.

Ma, che cosa significa $p < 0.05$?

Supponiamo di studiare un diuretico e prendiamo da una popolazione di 200 persone un campione di 10 persone (A) che verranno trattate con il diuretico e 10 persone (B) che faranno da controllo e trattate con placebo. Si tratta di calcolare l'incremento della diuresi dei due campioni e di calcolare se il diuretico è più efficace del placebo. Il ragionamento statistico si basa sull'ipotesi iniziale o "nulla" (H_0) che i due campioni appartengono alla stessa popolazione e che ciò che si osserva dipende dal caso. Quindi l'ipotesi iniziale H_0 è che la media della diuresi di A è = alla media della diuresi di B:

$$M_A = M_B \text{ oppure } M_A - M_B = 0$$

Se dal test t risulta che il diuretico è più efficace del placebo con probabilità statisticamente significativa con $p < 0.05$ significa che essendo i due campioni appartenenti alla stessa popolazione, l'unica differenza che li distingue è il trattamento con diuretico.

$p < 0.05$ è la probabilità di sbagliare nell'affermare che il diuretico è più efficace. In altre parole ho 5 probabilità su cento o una probabilità su 20 di sbagliare. Questa probabilità, però, è solo una convenzione arbitraria entrata nell'uso, ma ha il difetto di scoraggiare il ragionamento ed a sovrastimare un risultato.

Pericoloso è attribuire alla significatività statistica un'importanza clinica o una rilevanza biologica.

Uno studio potrebbe dimostrare che un farmaco antipertensivo è in grado di abbassare la pressione arteriosa di qualche mmHg ed essere altamente significativo, ma avrebbe scarsa importanza clinica.

Piccole differenze possono risultare statisticamente significative solo in considerazione dell'ampio numero del campione, mentre effetti rilevanti dal punto di vista clinico, possono risultare statisticamente non significative per lo scarso numero del campione.

Dopo queste premesse, alla domanda iniziale: "il farmaco A è migliore del farmaco B o del placebo" andrebbe reimpostata nel modo seguente: "di quanto il farmaco A è efficace rispetto al farmaco B"?

Alla domanda occorre dare una singola stima con l'aggiunta dell'indicazione dell'accuratezza della stima che viene espressa dall'intervallo di confidenza. Quest'ultimo esprime la probabilità di contenere il valore vero.

La probabilità comunemente usata è il 95%. Così l'intervallo di confidenza al 95% (esempio da 2 a 12) ci indica che abbiamo il 95% di probabilità di trovare il vero valore tra 2 e 12 e solo il 5% che non lo contenga.

Quindi l'intervallo di confidenza è un un test di significatività a tutti gli effetti ed inoltre ci dà molte più informazioni.

Facilita la distinzione tra significatività statistica e significatività clinica.

Facciamo un esempio pratico. Prendiamo un campione di 200 soggetti normali e controlliamo la pressione arteriosa sistolica e la confrontiamo con un numero uguale di soggetti diabetici.

Nei soggetti normali troviamo una PAS media di 140 mmHg e nei diabetici di 145 mmHg.

La differenza nella PAS è di 5 mmHg con intervalli di confidenza da 1 a 9 mmHg ed errore standard di 2 con $p < 0.05$. Quindi esiste significatività statistica tra i due campioni.

Da questi campioni noi estraiamo un certo numero di campioni provenienti dalla stessa popolazione e calcoliamo la PAS media e i relativi intervalli di confidenza al 95%.

Alla fine noteremo che il 95% degli intervalli di confidenza conterrà il vero valore della differenza della PAS media.

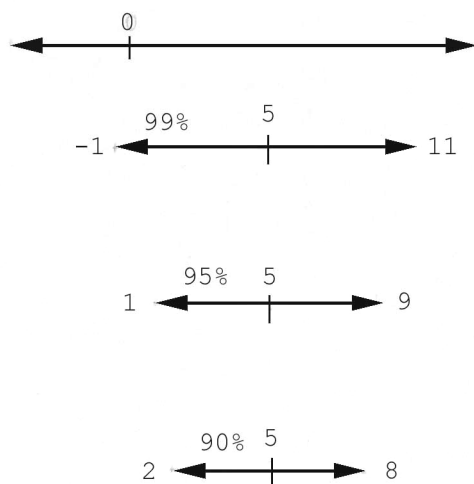
La dimensione del campione condiziona la dimensione dell'errore standard e quindi gli intervalli di confidenza.

(1) Più il campione è piccolo e più l'errore standard sarà grande e quindi ampi saranno gli intervalli di confidenza.

Se dal campione dei diabetici estraiamo un campione di 10 diabetici con PAS media sempre di 145 mmHg e calcoliamo gli intervalli di confidenza notiamo che questi ultimi diventano da -1 a 11 mmHg.

(2) Questo intervallo di confidenza contiene lo 0, pertanto ci dice che non esiste significatività statistica, mentre noi sappiamo che esiste perché l'abbiamo verificato in un campione grande di 200 diabetici. Questo fatto è sempre dovuto alla esiguità del numero campionario.

Un altro aspetto da considerare è il grado di confidenza. Normalmente si prende in considerazione il 95%. Come variano gli intervalli di confidenza al variare del grado di confidenza?



Maggiore è il grado di fiducia e maggiore sarà l'intervallo di confidenza

USO DELL'INTERVALLO DI CONFIDENZA

Negli studi gli IC vengono sempre usati in quanto forniscono una idea immediata della conclusione dello studio ovvero viene usato nella verifica delle ipotesi.

Infatti nel calcolo dell'IC entrano gli stessi parametri usati nel calcolo delle probabilità.

Per capire i valori dell'IC occorre prestare attenzione a 2 caratteristiche:

- **Se l'intervallo di confidenza contiene lo zero possiamo affermare che non ci sono prove sufficienti per rifiutare l'ipotesi di inefficacia. Questo in termini statistici. In parole povere e in termini comprensibili, non è efficace.**
- **Se l'intervallo di confidenza non contiene lo zero ci sono elementi sufficienti per rifiutare l'ipotesi di inefficacia, ovvero è efficace.**

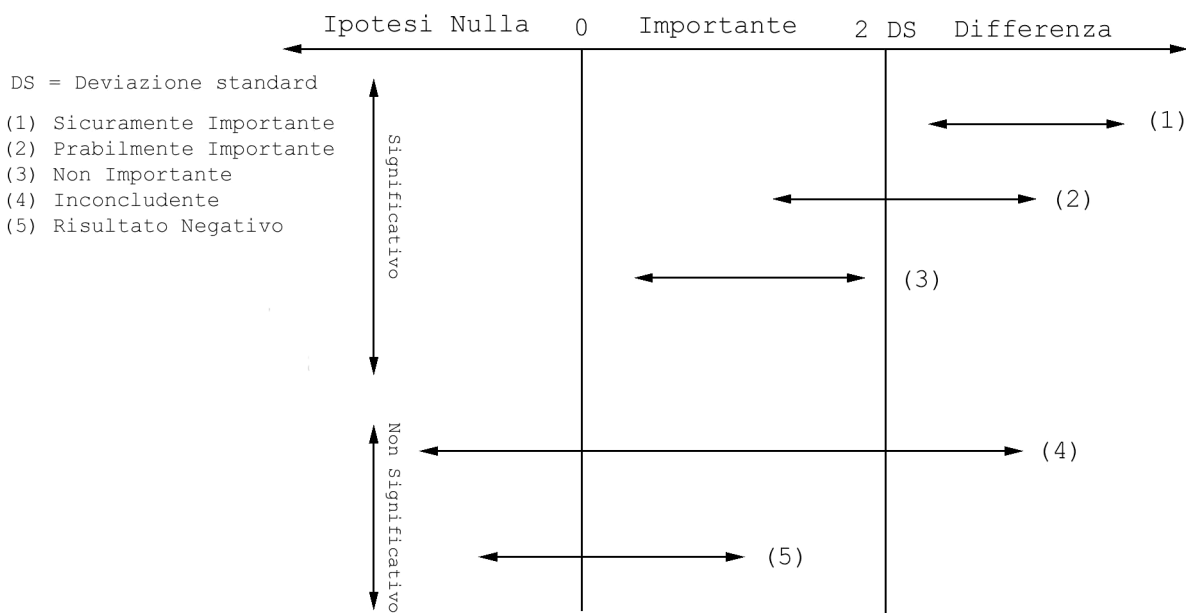
L'Intervallo di Confidenza presenta un altro vantaggio. Se un risultato raggiunge la significatività statistica, più a causa di un campione molto numeroso, che alla reale efficacia, l'IC lo mette in evidenza. Facciamo un esempio. Immaginiamo di trattare 100 persone con un antipertensivo e 100 persone con placebo. Il campione trattato con farmaco ha una diastolica media di 81 mmHg e una deviazione standard di 11 mmHg; il campione trattato con placebo ha una diastolica media di 85 mmHg. e una deviazione standard di 9 mmHg.

Tralasciando il calcolo dei dati di probabilità, si calcola, alla fine, una probabilità $P < 0.01$ che il farmaco abbassa la diastolica. Ma quanto, questo risultato, è significativo dal punto di vista clinico? Per rispondere alla domanda occorre calcolare l'intervallo di confidenza al 95%.

IC al 95% tra -6.8 e -1.2. In altre parole, il farmaco abbassa la diastolica tra 1.2 e 6.8 mmHg.

Quindi l'effetto è statisticamente significativo, ma è di entità clinica trascurabile. Il piccolo valore di p ($p < 0.01$) è dovuto più alla numerosità elevata del campione che all'effetto clinico.

La figura sotto illustra bene il concetto.



Potenza del test

Due tipi di errori.

Nel compiere un'analisi statistica possiamo incorrere in due tipi di errore:

- errore di tipo I in cui i dati affermano che esiste una differenza tra i campioni, quando in realtà non esiste (falso Positivo) e si indica con α . Convenzionalmente si indica 0.05 o 5% o 0.01 o 1% il massimo rischio accettabile.
- errore di tipo II quando i dati affermano che non esiste una differenza, mentre esiste (Falso negativo) e si indica con β .

La probabilità di rilevare un Vero Positivo si indica, quindi, con $1 - \beta$ ed esprime la potenza del test che non è altro che la sensibilità del test di rilevare un Vero Positivo.

In statistica la potenza del test cioè la sensibilità non dovrebbe mai scendere sotto l'80%, meglio se tra 80 e 90%.

Conclusione tratta dalle osservazioni	Situazione reale	
	Il trattamento è efficace	Il trattamento è inefficace
Il trattamento è efficace	Vero positivo, Conclusione corretta ($1 - \beta$)	Falso positivo, errore di tipo I (α)
Il trattamento è inefficace	Falso negativo, Errore di tipo II (β)	Vero negativo, conclusione corretta ($1 - \alpha$)

Va notato che i due tipi di errori sono interdipendenti. In pratica, quando si desiderano prove molto convincenti dell'efficacia di un farmaco si rende piccolo alfa, ma così si aumenta beta e quindi diminuisce la potenza del test ($1 - \beta$). L'unico modo per rendere piccolo sia alfa che beta è aumentare il numero del campione.

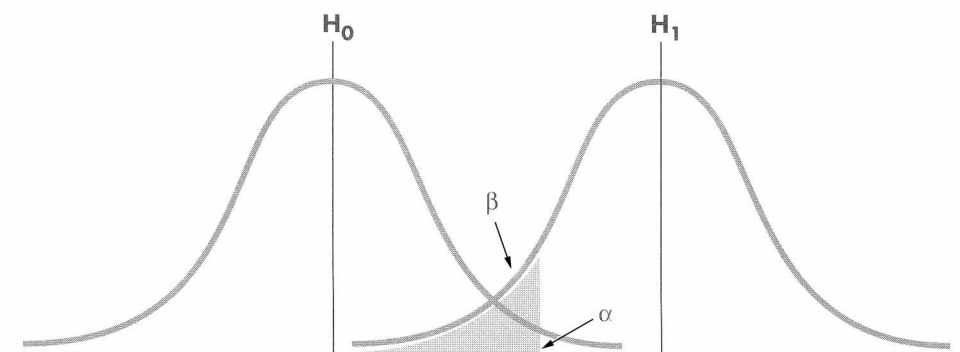
La potenza di un test dipende da tre fattori:

- 1) l'errore di tipo I che si vuole accettare nel rifiutare l'ipotesi di inefficacia del trattamento.
- 2) La differenza che si vuole rilevare
- 3) La numerosità campionaria

Esaminiamo i tre fattori separatamente.

1) La dimensione dell'errore di tipo I o alfa.

Alfa e beta sono complementari. Se diminuiamo alfa aumenta beta e di conseguenza diminuisce la potenza. La figura illustra la interconnessione tra alfa e beta.



Beta, rappresentato dalla area scura, rappresenta la zona di accettazione di H_0 (ipotesi nulla o assenza di differenza) che si sovrappone alla curva di H_1 (esiste differenza).

2) La differenza che si vuole rilevare.

La regola generale è che è più facile rilevare una differenza grande che piccola.

Più ampia è la differenza e più grande sarà la potenza. Se io volessi studiare un diuretico rispetto al placebo e misurassi l'incremento della diuresi con il diuretico rispetto al placebo, avrò una potenza maggiore se la differenza tra diuretico e placebo fosse di 400 ml/die rispetto a una differenza di 100 ml/die

Questa regola è facilmente comprensibile in quanto più grande è l'effetto del farmaco rispetto al placebo e più facile è affermare che ciò non sia dovuto al caso.

3) La numerosità campionaria.

Finora abbiamo stabilito che:

- la potenza decresce al crescere del grado di fiducia (da $P < 0.05$ a $P < 0.01$)
- la potenza cresce al crescere dell'entità dell'effetto cioè è più facile rilevare differenze grandi che piccole.

Sfortunatamente il ricercatore non può controllare questi due fattori.
L'unico sistema per controllare la potenza del test è aumentare la dimensione del campione.
La potenza cresce all'aumentare della numerosità campionaria per tre motivi:

- 1) Aumentando la numerosità campionaria cresce anche il numero dei gradi di libertà e il valore corrispondente all'errore alfa, che definisce l'errore di tipo I, decresce. (vedi tabella dei gradi di libertà)
- 2) Il valore di t (desumibile dalla sua formula e che comunque la sua spiegazione esula da questo libro) aumenta al crescere della dimensione n del campione.
- 3) Aumentando la numerosità campionaria diminuisce l'errore standard (ES) e quindi anche l'area di delle due curve H_0 e H_1 .

Fatte queste premesse, come possiamo calcolare la potenza del test?

1) Con deviazione standard σ conosciuta.

$$N > 2 \left[\frac{(z_{2\alpha} + z_{2\beta})\sigma}{\delta_1} \right]^2$$

Dove:

N è la numerosità del campione,

$z_{2\alpha}$ per un errore di tipo I di 0.05 è uguale a 1.96,

$z_{2\beta}$ per un errore di tipo II (β) al 95% corrisponde a 1.64, al 90% corrisponde a 1.282 e al 80% corrisponde a 0.842,

σ è la deviazione standard

δ_1 è la differenza media tra i due gruppi.

Esempio:

Vogliamo verificare due diuretici A e B sulla diuresi. Da precedente studio si sa che la deviazione standard della diuresi è di 0.5. Utilizzando un livello di significatività di 0.05 e una potenza dell'80% per una differenza media tra i due trattamenti di 0.25, quante persone dovranno essere utilizzate?

Applicando la formula sopra avrò:

$$N > 2 \left[\frac{(1.96 + 0.842) * (0.5)}{0.25} \right]^2$$

$$N = 62.8$$

Per ogni gruppo occorrono 63 persone.

Facciamo il caso inverso. Stesso esempio, ma quanto sarà la potenza se ho a disposizione 50 persone?

$$50 = 2 \left[\frac{(1.96 + z_{2\beta})(0.05)}{0.25} \right]^2$$

$$\frac{50}{2} = \left[\frac{(1.96 + z_{2\beta}) * 0.5}{0.25} \right]^2$$

$$\sqrt{\frac{50}{2}} = \frac{(1.96 + z_{2\beta}) * 0.5}{0.25}$$

$$5 * 0.25 = (1.96 + z_{2\beta}) * 0.5$$

$$\frac{1.25}{0.5} = 1.96 + z_{2\beta}$$

$$2.5 = 1.96 + z_{2\beta}$$

$$z_{2\beta} = 2.5 - 1.96$$

$$z_{2\beta} = 0.540$$

Dalla Tabella 1 $z_{2\beta}$ corrisponde a 0.2946. E' evidenziato in giallo.

Potenza = $1 - \beta = 1 - 0.2946 = 0.71$ ovvero 71%

2) La formula riportata ammette che si conosca la deviazione standard σ , ma nella realtà è raro.

Per ovviare a questo inconveniente si modifica la formula sopra nella seguente:

$$N > \left\{ \frac{z_{2\alpha} \sqrt{2\pi(1-\pi)} + z_{2\beta} \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right\}^2$$

dove

N è il numero del campione

$z_{2\alpha}$ per un errore di tipo I del 0.05 è uguale a 1.96,

$z_{2\beta}$ per un errore di tipo II (β) al 95% corrisponde a 1.64, al 90% corrisponde a 1.282 e al 80% corrisponde a 0.842,

π è la media delle due frequenze π_1 e π_2

π_1 è il tasso di successo del primo trattamento

π_2 è il tasso di successo del secondo trattamento

Esempio:

Si vuole controllare la frequenza di successo di un nuovo trattamento rispetto al convenzionale. La frequenza di successo del trattamento convenzionale è di 0.25. Se il nuovo trattamento aumentasse la frequenza di successo a 0.35, quanti pazienti includere nello studio con una potenza del test del 90% e un errore di tipo I del 0.05?

Usando la formula sopra e con $\pi=0.3$, $\pi_1=0.25$, $\pi_2=0.35$, $z_{2\alpha}=1.96$ e $z_{2\beta}=1.282$ si ha:

$$N = \left[\frac{1.96\sqrt{2*0.3(1-0.3)} + 1.282\sqrt{0.25(1-0.25) + 0.35(1-0.35)}}{0.1} \right]^2$$

$$N = \left[\frac{1.96\sqrt{0.42} + 1.282\sqrt{(0.1875 + 0.2275)}}{0.1} \right]^2$$

$$N = \left[\frac{1.96 * 0.648 + 1.282 * 0.6442}{0.1} \right]^2$$

$$N = \left[\frac{1.27 + 0.8259}{0.1} \right]^2$$

$$N = 439.3$$

A questo numero occorre aggiungere la correzione di Fleiss ($2 / (\pi_1 - \pi_2) = 2 / 0.1 = 20$) e si ottiene:

$$439.3 + 20 = 459.3$$

Lo stesso numero si può ottenere dalla Tabella 2. Il calcolo è evidenziato in giallo.

Se, invece, vogliamo calcolare la potenza con una data dimensione campionaria, dobbiamo usare le formule e non le tabelle perché imprecise.

Usando lo stesso esempio e avendo solo 300 pazienti e considerando un errore di tipo I del 0.05, quanto è la potenza dello studio?

Applicando la formula:

$$300 = \left[\frac{1.96\sqrt{0.42} + z_{2\beta}\sqrt{(0.1875 + 0.2275)}}{0.1} \right]^2$$

300 meno la correzione di Fleiss ($2 / \pi_1 - \pi_2$) cioè $2 / 0.1 = 280$

$$280 = \left(\frac{1.96 * 0.648 + z_{2\beta} * 0.6442}{0.1} \right)^2$$

$$\sqrt{280} = \frac{1.27 + z_{2\beta} * 0.6442}{0.1}$$

$$16.73 = \frac{1.27 + z_{2\beta} * 0.6442}{0.1}$$

$$16.73 * 0.1 = 1.27 + z_{2\beta} * 0.6442$$

$$1.673 = 1.27 + z_{2\beta} * 0.6442$$

$$1.673 - 1.27 = z_{2\beta} * 0.6442$$

$$z_{2\beta} = 0.403 / 0.6442 = 0.626$$

Usando la Tabella 1 (il calcolo è evidenziato in verde) 0.626 corrisponde a metà tra 0.62 e 0.63 e quindi ai numeri 0.2676 e 0.2643, la cui media è 0.26595. La potenza, $1 - 0.26595$, è uguale a 73,4%.

2) Studi caso-controllo.

In questi studi si utilizza l'odds ratio e la formula da utilizzare è la seguente:

$$p' = OR_1 * \frac{p}{p(OR_1) + 1 - p}$$

dove

OR_1 è la frequenza del rischio nei controlli che rappresentano la popolazione generale

p rappresenta l'entità del fattore di rischio

Esempio: si sa che un certo fattore di rischio abbia una incidenza del 20% nella popolazione generale e vogliamo sapere l'associazione, posto il rischio a 2, con una potenza dell'80% e errore di tipo I del 0.05. Quanti casi e controlli occorrono?

posto $p = 0.2$ e $OR_1 = 2$

$$p' = 2 \frac{0.2}{0.2 * 2 + 1 - 0.2}$$

$$= 0.3333$$

Usando la formula data sopra e cioè

$$N > \left\{ \frac{z_{2\alpha} \sqrt{2\pi(1-\pi)} + z_{2\beta} \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)}}{\pi_1 - \pi_2} \right\}^2$$

utilizzando π uguale a 0.2666 (0.3333 + 0.2 / 2), π_1 uguale a 0.3333, π_2 uguale a 0.2, $z_{2\beta}$ uguale a 0.842 (potenza dell'80%) e usando la correzione per la continuità di Fleiss troviamo

$$N = \left\{ \frac{1.96 \sqrt{2 * 0.2666(1 - 0.2666)} + 0.842 \sqrt{0.3333(1 - 0.3333) + 0.2(1 - 0.2)}}{0.3333 - 0.2} \right\}^2$$

$$N = \left\{ \frac{1.96 \sqrt{0.5332 * 0.7334} + 0.842 \sqrt{0.3333 * 0.6667 + 0.2 * 0.8}}{0.1333} \right\}^2$$

$$N = \left\{ \frac{1.96 \sqrt{0.391} + 0.842 \sqrt{0.2222 + 0.16}}{0.1333} \right\}^2$$

$$N = \left\{ \frac{1.96 * 0.6253 + 0.842 * 0.6182}{0.1333} \right\}^2$$

$$N = \left\{ \frac{1.2257 + 0.521}{0.1333} \right\}^2$$

$$N = 171.6$$

Usando la correzione di Fleiss

$$N = 171.6 + 2 / 0.1333 = 186.6$$

Lo studio dovrebbe comprendere 187 casi e altrettanti controlli. Lo stesso numero si trova usando la Tabella 3 (il calcolo è evidenziato in giallo).

In uno studio caso-controllo è bene avere uno stesso numero di casi e controlli. Ma se un gruppo contiene m persone e l'altro rm persone, allora lo studio è equivalente a uno studio con n persone in ogni gruppo dove:

$$\frac{2}{n} = \frac{1}{m} + \frac{1}{rm}$$

Cioè:

$$m = \frac{(r+1)n}{2r}$$

Esempio: supponiamo che nell'esempio sopra non si disponga di 187 casi. Di quanti casi avremmo bisogno se ci fossero due controlli per ogni caso?

Utilizzando la formula: $m = (r + 1)n / 2r$

Dove r è uguale a 2 (due controlli per ogni caso) abbiamo

$$m = \frac{(2 + 1) * 186.5}{2 * 2}$$

$$m = 139.9$$

Quindi 140 casi e 280 controlli.

LA RICERCA SU MEDLINE

Medline è il più fornito ed aggiornato archivio medico esistente al mondo, che viene costantemente ed incessantemente aggiornato con tutti i lavori più significativi ed autorevoli pubblicati dai vari centri di ricerca medica e scientifica in genere. Il database di Medline è divenuto talmente complesso e ricco che non è più gestibile con una semplice indicizzazione per argomenti; a tale scopo i lavori presenti sono quindi stati raggruppati in diversi sottoindici, selezionabili mediante chiavi di ricerca anche abbastanza complesse.

Esistono organizzazioni che forniscono un accesso gratuito a Medline, intendendo con questo la possibilità di effettuare una ricerca nell'oceano di abstract e lavori full-text disponibili. Se si riesce a reperire quanto richiesto, si può in un secondo tempo decidere di farsi spedire direttamente a casa il lavoro completo citato, in questo caso con il pagamento di una somma di denaro.

Limitiamoci a vedere le possibilità di ricerca che abbiamo a disposizione. Essenzialmente sono due: la ricerca diretta sul Web e la ricerca mediante interrogazione con e-mail.

La ricerca sul Web

In tal caso occorre conoscere l'URL di un'organizzazione che fornisca l'accesso alla banca dati di Medline. Una delle più efficienti è Healthgate, il cui sito si trova all'indirizzo:

<http://www.healthgate.com/HealthGate/MEDLINE/search.shtml>

Collegandosi con questa URL, ci si ritrova a dover digitare in una finestrella di input la propria chiave di ricerca. In tal caso possiamo scrivere una o più parole, direttamente (in inglese!), e selezionare alcuni parametri di ricerca: la lingua inglese di pubblicazione, la possibilità di avere l'abstract completo, la limitazione agli ultimi due anni di pubblicazione, la quantità di abstract da ricercare per ogni videata.

Esempio: ho bisogno di sapere se esistono su Medline lavori che trattino dell'efficacia della vaccinazione antinfluenzale. In tal caso scriverò nella casella di ricerca le parole *influenza vaccine effectiveness* separate da uno spazio, e premerò il pulsante *Search*. In pochi secondi, avrò visualizzato gli abstract che rispondono a questi requisiti, e, dopo averli letti, sarò in grado di decidere se ordinare o meno il testo intero di uno o più lavori. Sulle pagine presentate è infatti sempre presente l'opzione che permette di ordinare *full text* gli articoli che interessano, con pagamento alla consegna o mediante carta di credito.

La ricerca può anche essere maggiormente articolata, imponendo determinate condizioni positive o negative: in tal caso ci si dovrà servire degli operatori logici AND, OR e NOT, che, opportunamente inseriti tra le chiavi di ricerca, permettono di delimitare meglio i risultati ottenuti.

Se, ad esempio, volessi vedere i lavori pubblicati che riguardano i farmaci usati nella profilassi o nella terapia dell'asma bronchiale con l'esclusione dei cortisonici. In tal caso scriverò *Asthma therapy OR prevention NOT steroids*.

(Ovviamente il corsivo è usato in questa sede solamente per evidenziare meglio la scritta, e non deve essere adoperato)

La ricerca mediante e-mail

Una seconda possibilità di interrogazione dell'archivio di Medline è data dall'e-mail, in cui, senza passare attraverso un fornitore di servizi quale HealthGate, chiunque possenga un account ed un programma di e-mail può accedere ai database di Medline. In tal caso occorre inviare un messaggio all'indirizzo query@ncbi.nlm.nih.gov specificando nel corpo del messaggio i seguenti parametri (uno per riga):

DB
TERM
DOPT
DISPMAX

DB è la sigla di *database* e dopo le lettere DB occorre la lettera *m*, ad indicare che l'archivio che si vuole interrogare è quello di Medline. Occorre specificarlo in quanto con quest'indirizzo si possono anche interrogare, variando la lettera, i seguenti archivi:

n l'archivio dei nucleotidi

p l'archivio delle proteine

t l'archivio delle strutture molecolari in 3D

s l'archivio *n* e *p* contemporaneamente

Dopo la parola TERM occorre indicare la chiave di ricerca, esattamente come descritto prima per il Web.

Dopo la sigla DOPT (Display OPTion) conviene scrivere la lettera *r*, che si riferisce alla modalità di visualizzazione tipica dei record di Medline.

Dopo la sigla DISPMAX si può scrivere il numero massimo di record da visualizzare nella risposta.

Per esempio, una ricerca sull'efficacia della finasteride nell'alopecia sarà impostata in questo modo:

DB m

TERM alopecia AND finasteride effectiveness

DOPT r

DISPMAX 10

La risposta, salvo il caso di server particolarmente intasati, generalmente arriva nel giro di pochi minuti.

La creazione di meta-analisi

Una meta-analisi può anche essere creata per scopi personali, mettendo insieme più lavori clinici su argomenti che interessano. In tal caso conviene impostare e limitare la ricerca ai trials randomizzati e controllati (RCT) che sono disponibili, per avere un risultato più autorevole.

Esempio: supponiamo di voler fare una meta-analisi di tutti gli RCT riguardanti l'efficacia della determinazione del PSA quale screening del carcinoma della prostata.

In tal caso, le parole-chiave fondamentali che ci permettono di impostare la ricerca saranno semplicemente:

- RCTs
- PSA
- Screening

Potremo quindi inviare a query@ncbi.nlm.nih.gov un messaggio composto dalle seguenti righe:

DB m

TERM RCTs and PSA and screening

DOPT r

DISPMAX 50

Toccherà poi a noi vagliare il risultato di una simile richiesta, esaminare le caratteristiche dei trial presentati, estrapolarne i dati e riassumerli in una meta-analisi. Il discorso è comunque molto più complesso, dal momento che i dati devono essere resi omogenei tra loro mediante aggiustamenti delle variabili e dei fattori di rischio presenti; ma l'esempio suddetto è solamente un punto di partenza.