

Traduzione a cura di:
Infermiera DANIELA MOSCI
in collaborazione con D.D.S.I. Paolo Chiari
Centri studi EBN - Direzione Servizio Infermieristico e Tecnico
Azienda Ospedaliera di Bologna – Policlinico S.Orsola-Malpighi
Via Massarenti, 9
40138 Bologna, Italia

Tel. e Fax. 051 6363049

E-mail: ebn@orsola-malpighi.med.unibo.it

Web: http://www.med.unibo.it/reparti_servizi/servinfer/homepage.html

STATISTICA PER EVIDENCE BASED NURSING

I professionisti e gli amministratori sanitari sono sempre più consapevoli della necessità che le loro decisioni siano conformi alle migliori evidenze disponibili. Anche se gli articoli selezionati per la sintesi su Evidence-Based Nursing sono già stati valutati, e sono state scelte solo le ricerche di alta qualità, i professionisti hanno ancora necessità di sviluppare la propria abilità di identificare e ed interpretare le ricerche dal punto di vista metodologico. Questo editoriale è parte della serie finalizzata ad aiutarvi ad acquisire questa capacità, ed è il secondo dei tre editoriali che si focalizzano su alcune misure base e su problemi statistici che si trovano all'interno delle ricerche sanitarie. In questo editoriale, esamineremo come i risultati degli studi concernenti a trattamenti, o interventi, sono riassunti e presentati per differenti tipi di esiti sanitari. Il prossimo editoriale spiegherà come sono utilizzate le tecniche statistiche al fine di valutare la probabilità che gli effetti dei trattamenti osservati si verificano a causa del caso, e cosa si intende per risultato "statisticamente significativo".

LE MISURE DI SALUTE E DI MALATTIA

Sono utilizzate molte misure per valutare gli esiti sanitari di un intervento, da quelle che cercano di cogliere i loro effetti sulla salute della popolazione, fino a misure che valutano una dimensione specifica e rilevante per una particolare patologia. Alcune sono misure della percezione dei pazienti della loro patologia; più spesso, comunque, sono misure che i clinici o i ricercatori ritengono importanti. Queste misure possono essere sia continue che discrete; questa distinzione è importante perché il tipo di misura utilizzata determina il modo in cui i risultati sono presentati ed analizzati.

MISURE CONTINUE

Quando gli esiti degli studi sono continui (ad esempio la temperatura, la pressione arteriosa, la concentrazione di colesterolo), i ricercatori sono solitamente interessati all'intervallo di valori nel

quale essi si modificano a seguito dell'esposizione ad un intervento. Per confrontare gli studi che utilizzano misure di esiti continue possono essere comparati i valori medi delle variabili (ad esempio la media, la mediana) dopo il trattamento. In uno studio sul trattamento delle artriti, i ricercatori possono misurare le modifiche nel livello di dolore o di mobilità utilizzando una o più scale di valutazione della mobilità o del dolore. Per patologie che mettono in pericolo la vita, gli studi possono valutare gli esiti come la durata della sopravvivenza.

Recentemente, l'enfasi è stata spostata maggiormente su misure centrate sulla salute del paziente. Queste misure si basano sulla valutazione quantitativa compiuta dal paziente della sua condizione di salute, rispetto a diverse dimensioni, come la funzionalità fisica o sociale, la salute mentale, il grado di dolore, o l'insieme di questi punteggi considerati separatamente, in forma di un profilo di salute (ad esempio Nottingham Health Profile) o uniti prima in un singolo numero o indice (ad esempio Eurogol o il General Health Questionnaire). Tutte queste misure utilizzano scale di valori continui.

Una volta che il cambiamento del punteggio è stato determinato per ogni singolo paziente, le modificazioni complessive per il gruppo che ha ricevuto l'intervento sono confrontate con quelle del punteggio del gruppo di controllo. Come si fa? Molti fenomeni biologici, come l'altezza, sono distribuiti "normalmente" nella popolazione. Il termine normale, in questo contesto, si riferisce alla curva simmetrica a forma di campana che si delinea quando sono riportati i valori di un campione numeroso. La maggior parte di essi si raggruppa intorno ad un valore medio, mentre agli estremi si individuano pochi valori. Il punteggio di un gruppo di persone è solitamente riassunto utilizzando un valore medio – la media o la mediana. La media è calcolata sommando tutti i valori e dividendo il risultato ottenuto con il numero dei valori. La mediana è il valore che si colloca al centro della distribuzione dei valori; il 50% dei dati si colloca prima della mediana e il 50% si colloca oltre ad essa. Quando i dati sono

distribuiti normalmente, la media e la mediana sono uguali ed il valore medio è utilizzato per riassumere i dati; tuttavia, quando i dati non sono distribuiti normalmente, situazione solitamente conosciuta come “distribuzione deviata”, è più significativo utilizzare la mediana piuttosto che la media per sintetizzare i dati. Dato che non tutti i soggetti inseriti in un gruppo rispondono ad un intervento allo stesso modo, o con lo stesso grado, dobbiamo sapere qualcosa rispetto all'estensione nel quale i valori sono concentrati o dispersi. Ciò si può realizzare attraverso la descrizione della gamma di valori in un gruppo (dato dal valore minimo e dal valore massimo registrato). Altri metodi di registrazione di questa dispersione includono gli intervalli interquartili, entro i quali sta il 50% dei valori, o la deviazione standard che è la misura del valore medio di scostamento dalla media di tutti i valori individuali in quel gruppo; più bassa è la deviazione standard, più piccola è dispersione dei valori.

MISURE DISCRETE

Oltre alle misure degli esiti sanitari rappresentabili su scale continue, spesso gli studi si focalizzano su eventi sanitari discreti, come la frequenza di una patologia, la morte, le dimissioni ospedaliere. Queste misure si caratterizzano per il fatto che l'evento o si verifica o non si verifica; queste sono le misure discrete (anche dette dicotomiche).

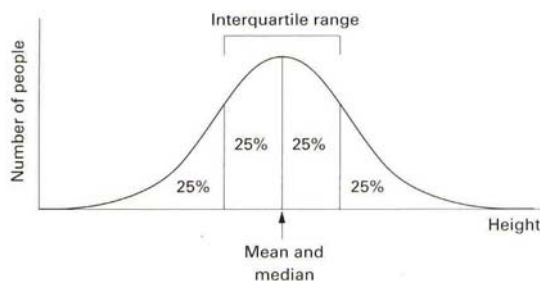


Figure 1 A normal distribution curve: variables such as height are distributed like this in the population.

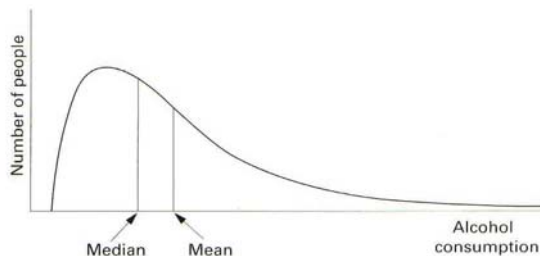


Figure 2 A skewed distribution curve: variables such as alcohol consumption are distributed like this.

Per un gruppo di persone, gli esiti discreti possono essere riassunti come la percentuale o proporzione di persone che fanno esperienza di un evento nel periodo di follow up. Per esempio, nello studio sulle fasciature compressive dopo l'angioplastica coronaria compiuta da Botti et al (vedi Evidence-

Based Nursing, luglio 1999, p. 84) il 3,5% dei pazienti con un bendaggio compressivo presentò sanguinamento, rispetto al 6,7% di coloro senza il bendaggio. Queste proporzioni esprimono la probabilità, o il rischio, che una persona nel gruppo di interesse presenti l'evento durante il periodo di follow up.

Questa misura riassuntiva può essere ampliata per considerare non solo se le persone fanno esperienza dell'evento, ma anche il tasso con il quale lo fanno. Per esempio se il 20% degli uomini in uno studio sono morti dopo 2 anni, il rischio degli uomini di morire al secondo anno sarà del 20% senza che si consideri se sono tutti morti durante il primo o il secondo anno. Il tasso, comunque, misura il numero delle persone che presentano l'evento nell'unità di tempo (incidenza). Se tutte le morti avvengono al primo anno, il tasso sarà 20 per 100 persone all'anno (20 morti per ogni 100 persone seguite per un anno), mentre se tutti muoiono al secondo anno, l'incidenza sarà dimezzata a 10 su 100 persone per anno.

Le misure continue sono spesso espresse come misure discrete negli studi valutativi, specialmente se c'è un limite sopra o sotto il quale vi sono differenze cliniche. Per esempio, le scale che misurano la depressione possono essere utilizzate per misurare il punteggio della depressione (cioè la stima della depressione) o per assegnare un valore al punteggio sopra il quale un paziente può avere una diagnosi di depressione. In altre parole, il primo approccio misura come le persone erano depresse prima e dopo il trattamento, mentre il secondo utilizza un valore soglia che classifica se più o meno persone erano depresse alla fine del trattamento rispetto a prima dello stesso. Un buon esempio del secondo approccio è dato nella analisi di St John's Wort di Linde e Murlow (Evidence-Based Nursing, luglio 1999, p. 82) dove i pazienti erano classificati come "responder" o "non-responder" al trattamento.

LE MISURE DEGLI EFFETTI E LE LORO ASSOCIAZIONI

La precedente sezione ha descritto i diversi modi di misurare gli esiti clinici. In questa sezione esamineremo come sono utilizzate queste misure per determinare se un intervento ha effetti sull'esito, l'entità di questo effetto e se è positivo o negativo. Come prima, l'approccio varia a seconda che l'esito sia misurato come una variabile continua o discreta. Ricordate che le differenze negli esiti dei pazienti, che ricevono interventi differenti non sono necessariamente determinati dall'intervento stesso. L'associazione non implica la causa. Può succedere che i pazienti nei gruppi siano stati selezionati in modo diverso, o che siano stati gestiti in modo differente per gli altri aspetti assistenziali, o che gli esiti siano stati selezionati in modo diverso. Questi sono criteri importati mediante i quali gli studi riguardanti i trattamenti devono essere valutati; questo argomento sarà trattato in un futuro editoriale

di Notebook. Vi possono essere anche differenze determinate dal caso, uno degli obiettivi delle analisi statistiche è determinare se il risultato conseguito può dipendere da esso.

MISURE CONTINUE

Negli studi che usano variabili continue per valutare gli esiti spesso vengono confrontati i valori medi (o mediani) degli esiti o il cambiamento medio negli esiti del gruppo sperimentale e del gruppo di controllo. Se i valori medi (o i cambiamenti) differiscono, ciò suggerisce che vi sono differenze negli effetti tra il trattamento sperimentato e quello di controllo. Per esempio, in uno studio riguardante il lutto negli omosessuali maschi (Evidence-based nursing, ottobre 1999, p. 116), i ricercatori hanno utilizzato un punteggio a struttura mista per il dolore e lo stress, ed hanno misurato i parametri prima e dopo il supporto per il lutto nel gruppo sperimentale, e prima e dopo nel gruppo di controllo che ha ricevuto una assistenza standard. I ricercatori hanno quindi confrontato i cambiamenti medi (in questo caso, una diminuzione) nel punteggio formato dalla valutazione del dolore e dello stress in ogni gruppo, ed hanno individuato una diminuzione significativa nel gruppo che ha ricevuto il trattamento sperimentale.

Un altro modo per esprimere le differenze negli outcome, per le misure continue, è mediante l'identificazione della variazione standard. Essa non ha unità di misura ed esprime semplicemente l'effetto di un intervento in termini del numero di deviazioni standard che si scostano dalle medie dei due gruppi. Questo permette il confronto della dimensione dell'effetto del trattamento quando sono utilizzate diverse misure degli outcome. Le deviazioni standard sono perciò utilizzate nelle meta-analisi, che sono finalizzate a confrontare ed a congiungere i risultati di molti diversi studi. Spesso incontrerete le variazioni standard all'interno delle revisioni sistematiche nella Cochrane Library. Tuttavia, dato che le deviazioni standard non hanno una unità predefinita, non hanno un diretto significato clinico e quindi sono difficili da interpretare.

MISURE DISCRETE

La misura degli effetti quando si utilizzano outcome discreti (ad esempio vivo o morto) pone a confronto il rischio di un evento nel gruppo sperimentale e nel gruppo di controllo. La tabella 1 lo illustra.

Il rischio di un evento nel gruppo di intervento è semplicemente la proporzione delle persone in quel gruppo che sperimentano l'evento,

$$R^i = \frac{a}{a + b}$$

Il rischio corrispondente nel gruppo di controllo è:

$$c$$

$$R^c = \frac{c}{c + d}$$

Questi sono due modi nei quali l'esito, o il tasso dell'evento, può essere confrontato nei partecipanti che hanno ricevuto o non hanno ricevuto l'intervento. Il rischio relativo, o tasso di rischio, è il rischio dei pazienti nel gruppo sperimentale di presentare l'esito in rapporto al rischio dei pazienti del gruppo sperimentale di presentare l'esito.

$$RR = \frac{R^i}{R^c} = \frac{\frac{a}{a + b}}{\frac{c}{c + d}}$$

Se il gruppo sperimentale e il gruppo di controllo hanno lo stesso esito, allora (assumendo che i gruppi siano comparabili in tutti gli altri aspetti), il rischio che si verifichi l'evento (ad esempio la morte) sarà lo stesso in entrambe i gruppi, e il RR sarà uguale a 1. Se il rischio di morire è ridotto nel gruppo sperimentale rispetto al gruppo di controllo, allora il rischio relativo sarà inferiore a 1. Tuttavia, se l'intervento è dannoso allora il RR sarà maggiore di 1. Tanto più RR si distanzia da 1, tanto più grande è la forza dell'associazione tra l'intervento e l'esito.

| | Evento Sì | Evento No | |
|---------------------|-----------|-----------|-------|
| Gruppo sperimentale | a | b | a + b |
| Gruppo di controllo | c | d | c + d |
| | a + c | b + d | |

Tabella 1 – numeri di persone che hanno ricevuto (o non ricevuto) un intervento e che hanno (o non hanno) un evento.

Per diverse ragioni statistiche, alcuni studi esprimono un outcome come probabilità dell'evento (a/b) piuttosto che come rischio dell'evento (a/a+b). L'odds ratio (OR) è la probabilità dell'evento nel gruppo sperimentale (a/b) diviso la probabilità che l'evento si presenti nel gruppo di controllo (c/d):

$$OR = \frac{a/b}{c/d} = \frac{ad}{bc}$$

Così come l'RR, un'OR pari a 1 significa che non vi sono differenze tra i gruppi, mentre un OR < 1.0 significa che l'evento è meno probabile all'interno del gruppo sperimentale, che nel gruppo di controllo.

Così, nello studio sugli infermieri divenute cliniche nella prevenzione secondaria (Evidence-based nursing, gennaio 1999, p. 21), vi fu una riduzione importante dei sintomi ed una riduzione della probabilità di ricovero ospedaliero, rispetto a quella riscontrata all'interno di programmi di cura standardizzati (OR = 0,64%); ossia, vi era una riduzione del 36% (1.0-0,64) della probabilità di ricovero ospedaliero. Quando l'evento misurato è abbastanza raro, l'OR e il RR sono numericamente simili perché i valori di a e di c sono insignificamente piccoli.

L'IMPORTANZA CLINICA

Gli OR e gli RR sono misure della forza dell'associazione tra intervento ed esito. Tuttavia, essi non ci dicono molto sull'impatto del significato clinico, e quindi se l'intervento è da considerare o meno.

È importante ricordare che il RR indica il beneficio relativo del trattamento, e non il beneficio effettivo; ciò significa che l'RR non considera il numero di persone che svilupperanno comunque l'esito negativo, misura che è possibile individuare con la riduzione del rischio assoluto - conosciuto come ARR quando il rischio è diminuito. L'ARR può essere calcolato semplicemente sottraendo la porzione di persone che hanno presentato l'evento avverso nel gruppo sperimentale, dalla porzione di persone che hanno avuto l'evento avverso nel gruppo di controllo. La differenza assoluta nel rischio ci dice quanto l'incidenza dell'evento avverso dipenda dall'intervento studiato.

Per esempio, in uno studio sulla depressione, se il 2% dei partecipanti del gruppo di controllo è depresso ($R^c = 0.02$) ed il trattamento riduce il rischio di depressione al 1% ($R^i = 0.01$), allora il rischio relativo è

$$\frac{R^i}{R^c} = \frac{0.01}{0.02} = 0.5$$

Un rischio relativo (RR) di 0.5 o una riduzione alla metà del rischio sembra essere un grande effetto, ma la riduzione assoluta del rischio è solo:

$$ARR = R^c - R^i = 0.02 - 0.01 = 0,01 \text{ o } 1\%$$

Questa è una riduzione del rischio di solo un evento su 100 persone, ed è molto meno impressionante che l'affermazione vista prima della riduzione della metà del rischio di depressione. Così, quando leggiamo un report di uno studio sperimentale, abbiamo bisogno di interpretare il RR o l'OR nell'ambito del contesto di quanto frequentemente dell'outcome ricorre nella popolazione.

Un'altra modalità di espressione dell'effetto di un intervento è il *number needed to treat* (NNT), che esprime in modo efficace l'effetto assoluto di un

intervento. Questo è calcolato semplicemente dividendo 1 per la riduzione assoluta del rischio:

$$NNT = \frac{1}{(R^c - R^i)} = \frac{1}{0.01} = 100$$

In altre parole, per prevenire un ulteriore caso di depressione occorre trattare 100 pazienti. Il NNT rappresenta il numero di pazienti che dovremmo trattare per prevenire un ulteriore evento, ed è un modo utile per esprimere l'efficacia clinica - maggiore è l'efficacia di un intervento, minore è il NNT.

Per illustrare questo concetto, considerate l'impatto dello stesso intervento, che dimezzi ugualmente il rischio di depressione nel gruppo di trattamento rispetto al gruppo di controllo, ma, questa volta, in una popolazione a rischio maggiore. Se il rischio di evento nel gruppo di controllo (R^c) è attualmente di 0,2 (cioè il 20% della popolazione presenta l'evento), allora il rischio nel gruppo sperimentale (R^i) sarà di =0,1 (cioè il 10% della popolazione presenta l'evento). Ora la riduzione assoluta del rischio è $0.2 - 0.1 = 0.1$, e il NNT è $1/0.1 = 10$. In altri termini, solo 10 persone dovranno essere trattate perché sia prevenuto un ulteriore caso di depressione. L'efficacia del trattamento in termini di OR e di RR è uguale, ma l'impatto è maggiore dato che il rischio di depressione nel gruppo non trattato, è molto più alta.

Un recente esempio di come la differenza tra il rischio assoluto ed il rischio relativo possono influenzare l'interpretazione di uno studio, è fornito dagli studi epidemiologici che confrontano il rischio di morte nelle donne che usano contraccettivi orali di seconda o di terza generazione¹. Questi studi riportano un OR pari a 1.5 per la morte nell'uso dei contraccettivi di terza generazione rispetto a quelli di seconda generazione (un aumento approssimativo del 50% del rischio di morte). Ciò sembra un risultato eclatante. Tuttavia il rischio di morte delle donne che assumono entrambe i tipi di contraccettivi orali è molto basso, ed aumenta per le pillole di terza generazione in termini assoluti rispetto a quelli di seconda generazione, di solo 6 morti per un milione di donne. In altre parole, si dovrebbe avere una morte extra per ogni 166.666 donne che assumono pillole di terza generazione.

Nel numero di aprile 2000, considereremo il concetto di tests statistici, significatività statistica, ed intervalli di confidenza.

TREVOR SHELDON, DSc
*Department of Health Studies,
 University of York,*

¹ Spitzer W, Lewis MA, Heinemann LAJ, *et al.* Third generation oral contraceptives and risk of venous thromboembolic disorders: an internazionale case-control study. *BMJ* 1996; 312:83-8.

York, UK

NOTA DELL'EDITORE

La riduzione assoluta e relativa del rischio può essere espressa in 4 modi differenti, in relazione all'outcome considerato (evento "positivo" o evento "negativo") e la direzione dell'effetto. Una riduzione

del rischio si verifica quando il rischio di evento negativo diminuisce. Un aumento dei benefici si verifica quando il rischio di un evento positivo aumenta. Un aumento del rischio si ha quando il rischio di eventi negativi aumenta, o quando si ha la riduzione dei benefici



Settembre 2002