

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

LA STIMA DEGLI EFFETTI DI TRATTAMENTO: SONO REALI O DIPENDONO DAL CASO?

Nel numero di Notebook del gennaio 2000 di Evidence-Based Nursing abbiamo descritto come possono essere misurati e riassunti i risultati degli studi clinici prima di essere analizzati. Ora discuteremo su come possiamo stabilire se, utilizzando test statistici, il trattamento ha un effetto reale sulla salute o se lo studio ha solo effetti apparenti determinati dal caso.

Quando leggiamo criticamente il report di un trial clinico, una delle cose che ci destano maggiore interesse è sapere se i risultati dello studio forniscono una stima accurata del reale effetto del trattamento, sul tipo di pazienti inclusi nello studio.

ERRORI DI CAMPIONAMENTO

Anche se uno studio è stato condotto in modo metodologicamente corretto (senza bias), un risultato come “il 5% delle lesioni guarite dal trattamento rispetto a quelle del gruppo di controllo” non significa necessariamente che vi sia una reale efficacia terapeutica. Questo risultato può essere stato determinato dal caso anche qualora non vi fossero stati effetti reali. Per comprendere ciò, immaginate di giocare con i dadi. Sappiamo che, in media, ognuno dei sei numeri dovrebbe uscire un numero uguale di volte, se il dado non è truccato. Tuttavia, qualora un vostro amico realizzi due o tre “6” consecutivamente, voi difficilmente (a seconda dell'amico) lo accusereste di avere truccato i dadi o di avervi ingannato. Probabilmente concludereste che ciò è accaduto esclusivamente per caso. Questo esempio ci dimostra che, anche se non ci sono effetti reali (cioè i dadi non sono truccati), potremmo osservare eventi che sembrano testimoniare la presenza di un effetto reale semplicemente a causa del caso (errore di campionamento). Questo caso si verifica soprattutto quando si hanno poche osservazioni. Per esempio, se il numero “6” esce in due lanci su quattro (cioè nel 50% delle volte),

possiamo dedurre che ciò è successo per caso. Invece, se il numero “6” uscisse cento volte su duecento lanci, allora tenderemmo a rifiutare l'ipotesi che ciò è accaduto solo per caso, ed potremmo considerare come spiegazione che il dado era truccato.

La stessa logica può essere applicata al risultato delle valutazioni degli interventi clinici. È possibile che uno studio dimostri un beneficio od un danno per caso, in modo particolare se si tratta di uno studio di piccole dimensioni. Quando analizziamo i risultati di uno studio, dobbiamo verificare i limiti entro i quali essi possono essere stati determinati dal caso. Se è molto improbabile che i risultati si siano determinati ad opera del caso, accetteremo che essi riflettono un effetto reale del trattamento.

Contrariamente a quanto avviene per il dado, che possiamo controllare ripetendo i lanci per verificare che il numero di “6” usciti è stato determinato dal caso, non possiamo ripetere gli studi clinici molte volte al fine di verificare se i risultati conseguiti sono reali. Dobbiamo accontentarci di eseguire un unico studio. Certamente, il ripetersi dei risultati degli studi è una parte importante dei processi scientifici, e saremmo più sicuri dei risultati se essi fossero confermati da molti studi.

La teoria statistica ci dice che se noi potessimo ripetere un esperimento centinaia di volte, con campioni differenti aventi lo stesso numero di pazienti, il risultato (ad esempio, la differenza media, la differenza in proporzione, o il rischio relativo) non sarebbe sempre lo stesso. Se noi disegnassimo i risultati di questi esperimenti su un grafico, la forma della curva (cioè la distribuzione dei risultati) sarebbe approssimativamente normale, o a forma di campana. In corrispondenza della media, i risultati degli studi ci darebbero una stima corretta degli effetti reali del trattamento. Tuttavia, qualsiasi risultato proveniente da uno studio può variare dall'effetto reale a causa del caso. Il grado di

diversità dall'effetto "reale" è dato dalla misura della dispersione, o deviazione standard, di questa distribuzione che è chiamato errore standard (ES), dato che indica l'entità della probabilità degli errori di campionamento. Tanto più grande è l'errore standard, più i risultati individuali si discosteranno dal valore reale.

INTERVALLI DI CONFIDENZA

Per nessuno studio possiamo essere certi di quale sia il valore reale del trattamento, poiché può variare dal valore reale per colpa del caso. Grazie della forma della curva di distribuzione del campione, sappiamo che il 95% di tutti gli studi possibili danno una stima che si pone a 1,96 dall'errore standard su ciascun lato del valore reale. Così possiamo dire che nel 95% degli esperimenti, il valore reale si collocherà entro 1,96 ES da ciascun lato della stima dell'effetto individuato nel nostro singolo studio. In altre parole, c'è il 95% di probabilità che questo intervallo (dato da 1,96 X ES rispetto ad entrambe i lati della misura degli effetti dello studio) includa il valore reale. Questo è chiamato intervallo di confidenza al 95%, ed è un plausibile intervallo entro il quale con la probabilità del 95% si trova il valore reale.

Se vogliamo essere più sicuri che il nostro intervallo di confidenza includa il valore reale, possiamo usare un intervallo di confidenza al 99% che si colloca nell'intervallo definito da 2,58 ES su ciascun lato del valore individuato dal nostro studio. In questo caso c'è solo una possibilità su 100 che il valore reale cada fuori dall'intervallo.

Più ampio è l'intervallo di confidenza, meno precisa è la stima dell'effetto del trattamento. Questa precisione dipende dalla grandezza dell'ES. Esso è una misura di dispersione della distribuzione del campione, e dipende dalla sua dimensione. Minore è il numero dei pazienti inclusi in uno studio, o degli eventi osservati (ad esempio le morti), maggiore sarà l'errore del campionamento. Maggiore è l'ES, maggiore è la probabilità che qualsiasi altro esperimento differisca, a causa del caso, dal valore reale o dal valore medio e conseguentemente maggiore sarà l'intervallo di confidenza al 95%. D'altro canto, se aumentiamo la grandezza dello studio in modo che la distribuzione divenga meno dispersa, i risultati degli studi individuali cadranno più vicino rispetto al valore medio o reale e l'errore standard e l'ampiezza dell'intervallo di confidenza sarà ridotto. Ciò vale a dire che siamo più sicuri dei risultati del lancio dei dadi se li tiriamo molte volte piuttosto che se li tiriamo poche volte.

Il metodo attualmente preferito da molte buone riviste di assistenza sanitaria per condurre la rappresentazione dell'incertezza degli effetti del trattamento è rappresentato dagli intervalli di confidenza. Essi possono essere usati anche per verificare se il trattamento ha effetto o meno. Se

l'intervallo di confidenza della differenza dalla media della riduzione della pressione arteriosa o la differenza nella percentuale di lesioni guarite include il valore zero (cioè non vi sono differenze nella media o nella percentuale), allora possiamo dire che non siamo certi che il trattamento presenti dei vantaggi rispetto al trattamento di controllo. Allo stesso modo, se l'intervallo di confidenza di un odds ratio o il rischio relativo includono il valore 1 (cioè l'indifferenza nella probabilità o nel rischio), non possiamo essere sicuri che ci sia una differenza tra gli effetti del trattamento sperimentale e del controllo. Ciò è equivalente a condurre un test delle ipotesi (vedi sotto).

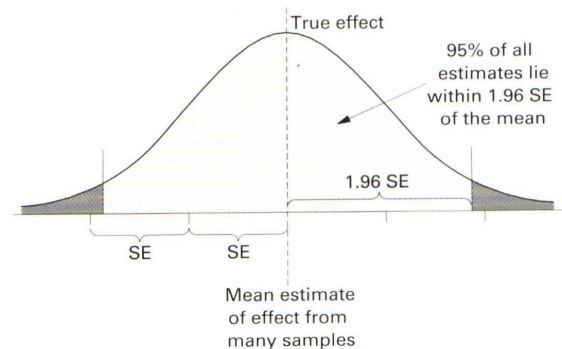


Figure 1 Sampling distribution showing effect of sampling error (SE).

Table 1 Larger sample sizes give more precise confidence intervals

Sample size (each group)	Estimated reduction in blood pressure	95% confidence interval
50	6 mm Hg	-1.0 to 13.0 mm Hg
100	6 mm Hg	1.1 to 10.9 mm Hg
200	6 mm Hg	2.5 to 9.5 mm Hg
1000	6 mm Hg	5.2 to 6.8 mm Hg

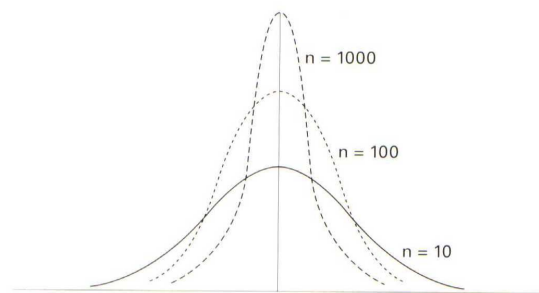


Figure 2 The larger the sample (n), the smaller the sampling error.

Possiamo illustrare quanto precedentemente detto con due esempi. Campbell ha valutato i vantaggi apportati dagli infermieri clinici nella assistenza primaria, rispetto ai miglioramenti condotti nella prevenzione secondaria delle patologie cardiache ed ha riportato un aumento dell'uso della aspirina (odds ratio 2,2, IC 95% 1,5-3,0)¹.

¹ Campbell NC, Mitchell LD, Thain J et al. Secondary prevention in coronary heart disease

Ciò significa che siamo sicuri al 95% che gli infermieri clinici determinino un aumento dell'assunzione di aspirina, da parte dei pazienti, in una misura compresa tra 2,15 volte fino al 0,8 volte rispetto agli standard assistenziali. Tuttavia, l'odds ratio per la cessazione del fumo individuato è di 0,8 (CI 95% 0,1-1,28). Ciò significa che il range di valori possibili include sia una riduzione della cessazione (la parte dell'intervallo compresa tra 0,8 a 1) ed un aumento della stessa (la parte compresa tra 1 e 1,28). Un odds ratio di 1 significa che gli infermieri clinici e il gruppo di infermieri standard hanno determinato lo stesso tasso di cessazioni (cioè non vi sono state differenze). Dato che l'intervallo di confidenza comprende il valore 1, non possiamo essere sicuri che la modifica degli interventi modifichi il tasso di cessazione dell'abitudine al fumo.

Un secondo esempio illustra l'uso degli intervalli di confidenza nel caso in cui la misura dell'effetto è data dalla differenza nella media di una variabile continua. Iitzman et al hanno valutato l'effetto delle visite domiciliari condotte dagli infermieri ad una popolazione di primipare a basso reddito² (vedi Evidence Based Nursing, luglio 1998, p. 6). Gli esiti hanno incluso il peso alla nascita, il numero di ospedalizzazioni ed il punteggio di stimolazione emozionale/cognitivo. Le medie, le differenze delle medie e gli intervalli di confidenza al 95% di queste differenze medie sono mostrate in tabella 2.

Variabili dipendenti	Media del gruppo visitato dalle infermiere	Media del gruppo di controllo	Differenza media	Intervalli di confidenza al 95%
Peso alla nascita in grammi	302	350	-18.2	Da -98 a 62
Punteggio di stimolazione cognitiva ed emozionale	2	0.9	1	Da 0 a 2.2

Tab. 1 Le visite domiciliari delle infermiere versus l'assistenza routinaria rispetto al peso alla nascita ed alla stimolazione cognitiva ed emozionale.

randomised controlled trial of nurse led clinics in primary care. Heart 1998;80:52.

² Iitzman, Olds DL, Anderson C, et al. Effect of prenatal and infancy home visitation by nurses on pregnancy outcome, childhood injuries, and repeat childbearing. A randomised controlled trial. JAMA 1992;268:52.

La differenza media nel peso alla nascita tra il gruppo esposto al trattamento ed il gruppo di controllo è di 18,2g. Tuttavia, gli intervalli di confidenza al 95% di questo valore hanno variato da -98,2g a 62,2g. Questa gamma di valori plausibili per la reale misura dell'effetto include una differenza uguale a zero, e quindi non possiamo inferire che vi siano degli effetti nelle visite domiciliari sul peso alla nascita. Possiamo tuttavia essere più sicuri della presenza di un reale miglioramento del grado al quale l'ambiente domestico risulta stimolante da un punto di vista mentale e cognitivo. L'incremento medio nel punteggio di 1,1 ha presentato intervalli di confidenza variabili da 0 a 2,2. Dato che questo intervallo non include lo zero, siamo certi al 95% che vi è stato un reale effetto del trattamento.

IL TEST DELLE IPOTESI ED IL VALORE DI p

Anziché cercare di valutare un plausibile intervallo di valori entro i quali il reale effetto del trattamento è probabile che sia (cioè l'intervallo di confidenza), i ricercatori spesso partono dall'assunto formale che non vi siano effetti (ipotesi nulla). Questo modo di ragionare è un po' simile alla situazione di un imputato davanti ad una corte in tribunale, nella quale si presume che egli sia innocente fino a prova contraria. Lo scopo della valutazione è simile a quello del procedimento penale: ottenere abbastanza prove necessarie a rifiutare l'ipotesi nulla ed accettare, invece, l'ipotesi alternativa, ossia che il trattamento ha un effetto (l'imputato è colpevole). Maggiore è la quantità e la qualità di evidenze che sono incompatibili con l'ipotesi nulla, maggiori sono le probabilità che rifiuteremo questa ipotesi ed accetteremo l'alternativa.

Nel caso della corte del tribunale, maggiori prove ci sono a carico dell'imputato, più probabilità vi sono che egli sia colpevole e che verrà condannato. In una valutazione clinica, maggiore è l'effetto del trattamento (espresso come la misura di errori standard diversi da zero), maggiore è la probabilità che l'ipotesi nulla non sia supportata e che accetteremo l'alternativa di una reale differenza tra il gruppo di trattamento ed il gruppo di controllo. In altre parole, il valore degli errori standard, che definisce se il risultato dello studio si discosta dal valore nullo, è equivalente, nella analogia del caso della corte di tribunale, all'ammontare delle prove contro l'innocenza dell'imputato. L'errore standard è considerato come l'unità di misura che indica se il risultato non dipende dal caso. Tanto più il risultato degli errori standard è lontano dal valore nullo, meno probabile è che esso si sia verificato per opera del caso, e maggiore è la probabilità che esso sia un effetto reale.

Per esempio, se lo studio mostra una differenza media nella pressione sanguigna di 5 mmHg, allora

il risultato è solo 1 errore standard sopra o sotto il valore nullo di indifferenza. Sappiamo che il 68% dei risultati di uno studio sono probabilmente collocati entro 1 ES da zero, anche quando non vi sono differenze di trattamento, semplicemente per opera del caso. In questa situazione, lo studio fornisce una evidenza insufficiente per rifiutare l'ipotesi nulla; il risultato è compatibile con la casualità.

Tuttavia, se il risultato dello studio è una differenza di 15,1 mmHg (cioè più di 3 errori standard sopra o sotto lo zero), allora è più facile rifiutare l'ipotesi nulla, dato che sappiamo che quando non ci sono reali effetti di trattamento, il 99,7% degli studi avrà un risultato compreso tra 3 errori standard e zero (15 mmHg). La probabilità che un esperimento dia come risultato una riduzione della pressione arteriosa di 15 mmHg per caso, quando non ci sono effetti del trattamento reali, è meno dello 0,3% (3 su 100).

Quale è secondo noi la probabilità (cioè che distanza deve esserci tra errori standard e zero) che possiamo assumere per dire che il risultato di uno studio non sia dovuto al caso, e che perciò esiste una reale differenza di trattamento? Abbiamo bisogno di definire alcune regole o criteri per decidere quando il risultato è da considerare compatibile con l'ipotesi nulla, e quando è così improbabile che sia determinato dal caso da farci rifiutare l'ipotesi nulla ed accettare invece, che il trattamento determina una differenza reale (l'imputato è colpevole nell'ipotesi del tribunale).

Per convenzione i ricercatori utilizzano probabilità <5% (1,96 ES) o < 1% (2,56 ES) come cutt-off. In altre parole, se l'effetto del trattamento cade oltre 1,96 ES sopra o sotto il valore nullo la probabilità che questo (o un risultato più estremo) si verifichi per effetto del caso, quando in realtà il valore nullo è vero, è <5% o 1 su 20. In questo caso diciamo che il valore dell'effetto del trattamento è statisticamente significativo al livello del 5% o con un valore di $p < 0,05$. Similmente, se il risultato è oltre 2,58 ES sopra o sotto lo zero, allora è considerato statisticamente significativo a livello dell'1% ($p < 0,01$). Per esempio nello studio sugli infermieri clinici, fu individuato un cambiamento medio nel punteggio complessivo di 6 componenti della prevenzione secondaria a oltre un anno di follow up di 0,59 nei pazienti assegnati agli infermieri clinici e 0,06 nel gruppo di controllo. La differenza tra i gruppi di 0,53 era oltremodo statisticamente significativa a $p < 0,001$; cioè, vi era meno di una probabilità di 1 su 1000 che il risultato fosse prodotto dal caso.

ERRORE DI TIPO I - IL RISCHIO DI RISULTATI FALSAMENTE POSITIVI

Ora ci dovrebbe essere chiaro che non possiamo sapere con completa certezza quale sia il reale effetto

del trattamento, anche quando non ci sono bias nello studio, determinati dal caso. Tutto quello che possiamo fare è dare una affermazione probabilistica, che indica con che probabilità il vero valore si colloca entro un range, e con che probabilità il risultato, o un risultato maggiore, è determinato dal caso. Ciò significa che c'è sempre la possibilità di essere in errore. Un tipo di errore è rifiutare l'ipotesi nulla in modo erroneo (cioè dire che l'effetto di un trattamento esiste quando in realtà il valore nullo è vero o, nel nostro esempio del tribunale, erroneamente imprigionare l'imputato). Questo falso positivo è chiamato *errore di tipo I*. Il rischio o la probabilità di questo tipo di errore è dato dal valore di p , o dalla significatività statistica dell'effetto del trattamento trovato. Se la differenza dell'esito prodotto dal trattamento è una riduzione della pressione arteriosa di 11 mmHg, con $p=0,04$, e noi inferiamo che esiste un reale effetto del trattamento, allora abbiamo un rischio di 4 su 100 di essere in errore, poiché il 4% di tutti gli esperimenti dello stesso tipo dovrebbero aver riportato una differenza uguale o maggiore a causa del caso. Più basso è il valore di p , e minore è la probabilità che vi sia un falso positivo e minore è il rischio di errore di tipo I. Ciò equivale a dire che tante più prove abbiamo, che supportano la colpevolezza dell'imputato, e minori probabilità vi sono di imprigionare un innocente. Il livello al quale l'errore di tipo I è accettato dipende dai costi e dalle conseguenze di realizzare una inferenza scorretta e di dire che un trattamento funziona quando in realtà non lo fa. Dato che ciò varia in relazione a ciò che è valutato, utilizzare un singolo cutt-off per tutti gli studi di ricerca non è molto saggio.

ERRORE DI TIPO II - RISCHIO DI UN RISULTATO FALSO NEGATIVO

Ovviamente, l'incertezza si realizza anche in altri sensi. Anche se in uno studio sono presenti prove insufficienti per dichiarare colpevole l'imputato - potrebbe accadere che egli sia rilasciato anche se colpevole. Analogamente, anche se non possiamo escludere la casualità come spiegazione del risultato dei nostri studi, ciò non necessariamente significa che il trattamento sia inefficace. Questo tipo di errore - un risultato falsamente negativo - in cui abbiamo erroneamente accettato le ipotesi nulle, relative alla assenza di effetti del trattamento, è detto errore di tipo II.

Maggiore è l'ampiezza della distribuzione dei valori (ES più ampi), più difficile è escludere la casualità, e quindi, maggiore è la probabilità di accettare erroneamente ipotesi nulle relative alla mancanza di effetti terapeutici del trattamento. Questo è un problema particolarmente presente negli studi con piccoli campioni, poiché essi hanno ampi ES. In questi casi, anche ampie stime dell'effetto del

trattamento non forniscono una sufficiente evidenza di reale efficacia del trattamento (cioè, non sono statisticamente significative) perché l'errore standard è così ampio. Possiamo dire che lo studio ha un ridotto potere di individuare l'effetto del trattamento statisticamente significativo, quando realmente c'è.

Se lo studio è troppo piccolo, gli intervalli di confidenza possono essere così ampi da non potere realmente escludere dall'intervallo il valore che indica la presenza di nessun effetto. Per esempio, molti studi sugli agenti, che compiono l'asportazione del tessuto necrotico nel trattamento delle lesioni croniche, sono così piccoli che individuano intervalli di confidenza molto ampi. Uno studio sul cadexomer iodine, per esempio, ha riportato un odds ratio per la guarigione delle lesioni di 5,5 a favore del cadexomer iodine rispetto al dextranomer³. Tuttavia, dato che solo 27 pazienti erano inclusi nello studio, l'intervallo di confidenza individuato era molto ampio, variabile da 0,88 fino a 34,48. Questo studio era troppo piccolo per essere in grado di escludere un odds ratio di 1 (nessuna differenza di trattamento).

Quando viene intrapreso uno studio, il numero dei pazienti dovrebbe essere tale per consentire di avere sufficiente capacità per rifiutare l'ipotesi nulla se esiste un effetto terapeutico clinicamente importante. Nel momento in cui progettano uno studio, i ricercatori dovrebbero condurre un calcolo che determini la capacità o la dimensione del campione in modo che abbia ragionevoli possibilità di rifiutare a giusta ragione l'ipotesi nulla. Questo importante calcolo dovrebbe essere riportato nell'articolo.

Uno dei problemi nelle ricerche cliniche è il vasto insieme degli studi troppo piccoli e che quindi hanno una capacità insufficiente. In questi casi, non si può interpretare un risultato come statisticamente non significativo per dire che non esistono effetti terapeutici.

Dato che molti studi sono troppo piccoli ed hanno poco potere, è possibile che siano persi importanti effetti clinici. Un approccio per risolvere questo problema è unire o combinare i risultati di studi simili, per ottenere una stima complessiva e più precisa degli effetti del trattamento. Questo approccio è detto meta-analisi e verrà discusso in un prossimo Notebook.

TEST PER DIFFERENTI TIPI DI MISURA DI ESITO

Nel numero di gennaio di Evidence-Based Nursing, abbiamo descritto diversi modi di esprimere gli effetti dei trattamenti, in relazione al tipo di misura

³ Tarvainen K. Cadexomer iodine (Iodosorb) compared with dextranomer (Debrisan) in the treatment of chronic leg ulcers. Acta Chir Scand Suppl 1998;544:57-9.

di outcome utilizzata. Questi comprendono anche il tipo di test statistici utilizzati per determinare l'estensione nella quale la stima dell'effetto del trattamento è determinato dal caso.

MISURE CONTINUE

Quando uno studio utilizza misure continue, come la pressione arteriosa, l'effetto del trattamento è spesso calcolato misurando le differenze del miglioramento medio dei valori pressori nei due gruppi. In questi casi (se i dati sono distribuiti in modo normale), è normalmente utilizzato un "t-test". Tuttavia, se i dati sono devianti (cioè, non distribuiti normalmente), è meglio verificare le differenze nella media utilizzando dei test non parametrici, come il test "Mann Whithney U".

CATEGORIE DI VARIABILI

Quando uno studio misura categorie di variabili ed esprime i risultati come proporzioni (ad esempio numero di lesioni infette o guarite), allora è utilizzato il test del "chi quadro". Questo verifica l'estensione nella quale la differenza tra la proporzione osservata nel gruppo del trattamento è diverso da ciò che ci si sarebbe aspettato dal caso se non ci fossero reali differenze tra il trattamento ed il gruppo di controllo. In alternativa, se è utilizzato l'odds ratio, può essere calcolato l'errore standard dell'odds ratio e ipotizzando una distribuzione normale, calcolando gli intervalli di confidenza al 95% ed il test delle ipotesi.

ANALISI ACCOPPIATE

I test descritti sopra si applicano a situazioni in cui sono confrontati gruppi indipendenti di pazienti. Ci sono, tuttavia, situazioni in cui i pazienti sono messi a confronto, o dove i pazienti sono utilizzati come controllori di sé stessi. In questi confronti a coppie, non è appropriato utilizzare i test illustrati sopra e sono necessarie analisi accoppiate. Per le misure continue normalmente distribuite, si può utilizzare il "t-test accoppiato". Per variabili non continue, si può utilizzare il "Wilcoxon signed-rank" test. Nel caso di categorie di outcome, sono utilizzabili diversi test, come il test di "McNemar". La cosa importante da ricordare è che se il progetto del confronto prevede l'accoppiamento o il confronto, così devono essere le analisi.

LA SIGNIFICATIVITÀ STATISTICA NON È SIGNIFICATIVITÀ CLINICA

Fino ad oggi, ci siamo concentrati sui modi di verificare se le stime dell'effetto del trattamento individuate in uno studio sono accurate e plausibili rispetto all'effetto reale, o se esse sono probabilmente dovute al caso.

I ricercatori e i lettori delle ricerche spesso focalizzano la loro attenzione sulla verifica che il

risultato sia statisticamente significativo (cioè, probabilmente non dovuto al caso). Tuttavia, proprio perché un test dimostra il risultato di un trattamento statisticamente significativo, non significa che il risultato sia clinicamente importante. Per esempio se uno studio è molto ampio (e quindi ha un piccolo errore standard), è più facile trovare effetti di trattamento piccoli e poco importanti che sono statisticamente significativi. Un grande studio randomizzato e controllato ha confrontato la riospedalizzazione nei pazienti che assumono un nuovo farmaco per il cuore con i pazienti che ricevono assistenza routinaria. Questo studio ha riportato una riduzione della riospedalizzazione dell'1% nel gruppo di trattamento (49% riospedalizzati rispetto al 50% del gruppo di controllo assistito in modo usuale). Ciò era statisticamente molto significativo ($p < 0,0001$), principalmente perché si trattava di uno studio ampio. Tuttavia, è improbabile che la pratica cambierà sulla base di una così piccola riduzione della ospedalizzazione.

RIASSUNTO

Quando leggiamo il report di uno studio clinico, è

importante ricordare che le differenze negli esiti dei pazienti nei diversi gruppi di trattamento possono derivare dal caso e non necessariamente da reali effetti di trattamento. Idealmente i risultati degli studi clinici (ad esempio le differenze di rischio, odds ratio, differenze dei valori medi) sono presentati con intervalli di confidenza intorno ad essi.

Gli intervalli di confidenza rappresentano il grado di incertezza intorno al risultato: tanto più stretti sono gli intervalli di confidenza, tanto più preciso è il risultato. Quando è presente una differenza statisticamente significativa, è importante anche considerare se la differenza è clinicamente importante e sufficientemente ampia da garantire cambiamenti della pratica.

Bibliografia: vedi articolo originale

TREVOR A SHELDON, DSc
Department of Health Studies
University of York,
York, UK



Settembre 2002