

Traduzione a cura di:
Infermiera MARIA CRISTINA ROBB
Centro studi EBN - Direzione Servizio Infermieristico e Tecnico
Azienda Ospedaliera di Bologna - Policlinico S.Orsola-Malpighi
Via Massarenti, 9
40138 Bologna, Italia

VALUTAZIONE DI STUDI SUGLI STRUMENTI DI SCREENING E DI VALUTAZIONE E SUI TEST DIAGNOSTICI

Nell'assistenza infermieristica la valutazione diagnostica è generalmente un'attività regolata e attuata dalle infermiere e dalle ostetriche. La valutazione e lo screening dei pazienti per ulteriori analisi (accertamento del caso), sono, comunque, elementi centrali dell'assistenza infermieristica. Il numero degli studi che esaminano i test e gli accertamenti è in aumento, ma la qualità metodologica di questi studi è, nell'insieme, scarsa. Quindi, le infermiere dovrebbero essere in grado di valutare in modo critico le evidenze di ciascuno studio per assicurarsi che siano stati usati gli strumenti di valutazione e di screening di qualità migliore. Gli strumenti di valutazione, accertamento del caso (o screening) e di diagnosi sono valutati usando criteri diversi da quelli applicati agli studi che valutano interventi terapeutici o di prevenzione, sebbene le tre domande di base per la valutazione critica siano le stesse: i risultati sono validi? Quali sono i risultati? I risultati mi aiuteranno nell'assistenza ai miei pazienti?

In questo articolo descriveremo, a grandi linee, un lavoro di critica agli studi che valutano uno strumento di screening per l'accertamento dei pazienti depressi, di Sackett e al.. Lo stesso lavoro può essere applicato anche agli studi sugli strumenti di valutazione, quali l'assegnazione di un punteggio per la valutazione del rischio di cadute o del rischio di piaghe da decubito, o agli studi sui test diagnostici.

SCENARIO CLINICO

Voi siete un infermiere che opera sul territorio e prestate assistenza, al suo domicilio, ad un uomo di 68 anni, con un'ulcera da diabete. Egli ha la sensazione che la sua ulcera ci stà mettendo un tempo lunghissimo per guarire e che non starà mai più bene di nuovo. Voi sapete, da precedenti conversazioni, che sua moglie è morta alcuni anni fa e i suoi due figli vivono fuori zona o all'estero. Quando poteva guidare era socialmente attivo, ma da quando è diventato dipendente da altri che lo assistono, non esce più molto. Egli riferisce che la sua alimentazione è buona e che la concentrazione di glucosio, con la sua terapia ipoglicemica, si sta stabilizzando in un range di normalità. Egli dice

anche che si sente sempre stanco. Voi avete notato che, recentemente, non si sta prendendo molta cura del suo aspetto esteriore e che sembra molto meno interessato agli eventi del mondo rispetto a quando avete cominciato a medicare la sua ulcera. Sebbene voi sappiate che il senso di malessere è un effetto collaterale di alcuni ipoglicemizzanti, il suo regime terapeutico non è stato modificato di recente. Basandosi su indagini successive, sembra improbabile che il paziente sia anemico, così iniziate a considerare l'eventualità della depressione. Voi sperate che vi siano dei modi semplici per la valutazione che potrebbero aiutarvi a sottoporre i pazienti ad ulteriori esami appropriati.

LA RICERCA

A differenza degli studi sugli interventi di prevenzione o terapeutici, che sono esaminati nel modo migliore dagli RCT o dalle revisioni sistematiche di questi lavori, i problemi relativi all'efficacia degli strumenti di valutazione e di screening nella pratica clinica, sono affrontati in modo migliore da studi incrociati su un gruppo. Gli studi sui test diagnostici e gli strumenti di screening possono essere un po' difficili da localizzare. La frase chiave sensibilità e specificità è utile, ma non sempre gli studi possono essere sempre localizzati con essa. In questo caso, quindi, può essere utilizzata la sigla /di (che sta per diagnosi). I database specialistici come la Cochrane Library non includono studi sui test diagnostici, quindi, dovete eseguire la ricerca seguente su Medline (OVID):

*Depression/di

*Depression disorder/di

Questionnaires/

(1 OR 2) AND 3

Vengono, così, identificati 303 articoli. Aggiungendo la parola di testo primary care, e limitando la ricerca ai lavori in lingua inglese, con l'abstract pubblicato negli ultimi 5 anni, si produce una lista di 24 citazioni più gestibile, tra queste vi è uno studio sugli strumenti di screening che include un semplice strumento, composto da 2 domande ("Durante il mese passato siete stato spesso turbato dalla sensazione di sentirvi giù, di essere depresso o senza

speranze?” e “ Durante il mese appena passato vi siete sentito senza interesse o piacere nel fare le cose?”). Whooley e al. hanno testato 7 strumenti di screening, su 542 pazienti consecutivi, che si presentavano in una clinica per l’assistenza urgente. Il 97% dei partecipanti erano uomini, la maggioranza dei quali era disoccupato. La prevalenza della depressione era del 18%. Tutti gli strumenti di screening riportavano risultati simili, ma gli investigatori raccomandavano, per la sua semplicità, lo strumento a due domande, da utilizzare nell’assistenza primaria. Prima di accettare le conclusioni degli autori, i lettori hanno bisogno di assicurarsi che lo studio sia valido. Questo richiede di rispondere a 4 quesiti:

**I RISULTATI SONO VALIDI?
C’È UN CONFRONTO CON UNO STANDARD
DI RIFERIMENTO DI DIAGNOSI (GOLD),
INDIPENDENTE E CIECO?**

Ci sono 2 aspetti in questo quesito. Primariamente, l’accuratezza di qualsiasi strumento è determinata, in modo migliore, comparando i suoi risultati con quelli ottenuti da un test di riferimento comunemente accettato. Questo viene spesso denominato il gold standard ed è spesso più invasivo del test iniziale. In questo modo la palpazione della fronte di un bambino con la febbre, potrebbe essere confrontata con la lettura di un termometro a mercurio per ottenere una stima veritiera del fatto che un bambino abbia una febbre più alta. Nello stesso modo, l’indice caviglio-brachiale della pressione (ABPI), potrebbe essere comparato con il gold standard di una venografia, per valutare se si testa l’ABPI come un segno di malattia arteriosa della gamba. I lettori hanno bisogno di essere sicuri che il test di riferimento sia il gold standard per la malattia. Confrontando la palpazione con una misura meno accurata di un termometro a mercurio (come un termometro timpanico), si potrebbe pervenire ad una stima poco accurata di quanti pazienti hanno realmente la febbre. Anche se lo studio utilizza il termometro a mercurio o uno strumento di similare accuratezza, il lettore deve essere sicuro anche che sia stata applicata una tecnica accettabile di termometria. Cioè, anche se è stato usato il termometro a mercurio, è più probabile che la metodica ascellare, fornisca una stima inattendibile della temperatura. Se partecipano più di un valutatore, lo studio dovrebbe fornire anche una stima del livello di consenso tra i valutatori.

Il secondo aspetto del quesito mette in guardia contro i bias di aspettativa. Sebbene nella maggior parte delle situazioni cliniche, gli operatori sanitari hanno accesso alle registrazioni dei pazienti, è importante, quando si valuta uno strumento, che i clinici si formino le proprie idee sulla malattia del paziente. La conoscenza precedente della presenza o dell’assenza di una malattia, potrebbe influenzare la valutazione di un clinico. Pertanto, è imperativo che i clinici che

eseguono una valutazione, usando il test gold standard, siano separati da quelli che usano l’altro strumento e che i 2 gruppi di clinici non siano a conoscenza della valutazione degli altri. Uno studio metodologico di valutazione di test diagnostici ha evidenziato che le valutazioni non cieche sovrastimano le diagnosi corrette di più del 30%, rispetto agli studi ciechi.

**IL TEST DIAGNOSTICO E’ STATO
VALUTATO IN UNA GAMMA APPROPRIATA
DI PAZIENTI (COME QUELLI CHE
INCONTRIAMO NELLA PRATICA CLINICA)?**

L’impresa più ardua, quando si valutano gli strumenti di accertamento del caso o di screening, è la loro applicazione alla popolazione considerata. I test derivano spesso da metodi “veloci e sporchi”, utilizzando una popolazione accessibile di pazienti, noti per avere la malattia in esame, e un gruppo di controllo sano. Se uno strumento non riesce a discriminare tra quelli con e senza la malattia, a questo stadio di sviluppo, è improbabile siano clinicamente utili. Ma, l’importanza di una valutazione può ingannare sulla sua abilità nel distinguere l’intera gamma degli attuali pazienti con la malattia (così come quelli che presentano sintomi simili, derivanti da malattie diverse), da quelli che non hanno la malattia. Dal punto di vista diagnostico, è più semplice distinguere i pazienti con una sintomatologia palese da quelli senza la malattia, piuttosto che identificare quelli con una sintomatologia sfumata. Solo se lo strumento è in grado di differenziare quelli che hanno maggiori probabilità di presentare la malattia da quelli che non costituiscono una reale popolazione clinica, può dimostrarsi utile. Gli esami di nuovi test spesso omettono lo stadio di sviluppo fondamentale della valutazione in una vera popolazione clinica. Per esempio, si può utilizzare l’ABPI nella valutazione dei pazienti con ulcere alle gambe, per evidenziare quelli con malattie arteriose periferiche, che dovrebbe essere esclusi dalla terapia con bendaggio compressivo. Negli anni sessanta, il valore normale di un ABPI, fu stabilito testando 110 pazienti che si conosceva avessero una malattia arteriosa oclusiva e confrontando i valori dei loro test con quelli di 25 controlli sani. Solo recentemente l’utilità dell’ABPI è stato testato in una popolazione territoriale simile a quella in cui viene comunemente usato dalle infermiere di distretto.

Generalmente viene accettato che sia buona pratica per gli studi di includere pazienti consecutivi che sono d’accordo di partecipare (minimizzando i bias di selezione potenziali), sebbene l’inclusione non consecutiva non sia stata trovata che abbia nessun effetto significativo sui risultati dello studio.

**LO STANDARD DI RIFERIMENTO È
APPLICATO SENZA TENER CONTO DEI
RISULTATI DEL TEST?**

Per evitare bias di verifica o di elaborazione, è necessario somministrare ai partecipanti entrambi i test, senza tener conto del risultato del primo test. Se il primo test è negativo e i partecipanti non ricevono il test gold standard per verificare questo risultato, i risultati dello studio saranno distorti. In alcuni casi, i partecipanti che hanno avuto un risultato negativo al test possono decidere di non sottoporsi al test gold standard, specialmente se questo è una procedura invasiva come la venografia. Piuttosto che escludere questi partecipanti, gli investigatori possono seguirli per un periodo appropriato di tempo e monitorare i sintomi della malattia in esame.

IL TEST E' STATO VALIDATO IN UN SECONDO GRUPPO DI PAZIENTI?

Per un lettore, per essere sicuro che i risultati dello studio sono accurati e non il risultato di idiosincrasie della coorte iniziale dei partecipanti o dell'abilità individuale dei valutatori, lo strumento dovrebbe essere valutato in un secondo gruppo indipendente di pazienti. Se i risultati sono i medesimi, gli operatori sanitari possono essere più sicuri sull'accuratezza dei risultati del test. Per esempio, l'uso combinato della colorazione di gram e il test all'Aridine-Orange Leucocyte Cytospin per diagnosticare rapidamente le infezioni sistemiche correlate a cateterismo, senza la rimozione del presidio venoso, è definito in modo favorevole, ma lo studio non valuta il test su un secondo gruppo di pazienti, da ciò la necessità di studi di conferma prima che il test sia largamente accettato.

PER RISPONDERE AL QUESITO ORIGINALE

Lo studio di Whooley e al. probabilmente incontra 3 criteri di validità su 4. Per prima cosa, gli investigatori confrontano gli strumenti di accertamento del caso con uno standard di riferimento accettabile, una versione computerizzata della scheda di intervista diagnostica (DIS). Questa è un'intervista di 20 minuti, con una sensibilità dell'80% e una specificità dell'84%, rispetto ai criteri DSM III per la depressione. Tre intervistatori preparati e all'oscuro del risultato derivato dagli strumenti di screening, somministrano il test di riferimento. Il livello di consenso interrelato tra i tre intervistatori, rispetto ai risultati del test di riferimento ($\kappa=0,88$), è stato alto. In secondo luogo, il campione dello studio era rappresentato da una popolazione realmente clinica, in una situazione di assistenza primaria, con pazienti che presentavano l'intera gamma delle rappresentazioni di depressione: episodi recenti di depressione, una depressione che dura da tutta la vita e l'assenza di depressione. Bisogna fare delle considerazioni caute su alcune delle caratteristiche del campione (come l'alta prevalenza di depressione, la suddivisione tra uomini

e donne, e l'alto numero di disoccupati), ma queste possono essere esaminate quando si considererà l'applicabilità dello studio ai vostri pazienti. In terzo luogo, tutti sette gli strumenti di screening e il test di riferimento, sono somministrati a 542 partecipanti consecutivi che accedono ad una clinica per cure urgenti di un centro medico, sebbene i risultati di 7 partecipanti siano stati esclusi dall'analisi per la perdita dei dati. I risultati degli strumenti di accertamento del caso non appaiono essere influenzati se viene somministrato il DIS. Comunque, lo studio non risponde al quarto criterio di validità, in quanto i risultati dello studio non sono stati valutati in un secondo gruppo. Nessun'altra valutazione dello strumento sembra essere stata fatta, sebbene una sia attualmente in corso in una popolazione di base in Nuova Zelanda (comunicazione personale B Arroll).

QUALI SONO I RISULTATI?

Quando i pazienti si presentano agli operatori sanitari, è probabile abbiano particolari malattie. Questa probabilità è la prevalenza di base di ciascuna malattia sul territorio. Ma ogni paziente è diverso. Prendiamo in considerazione due pazienti, che presentino entrambi una piccola ulcera del malleolo mediale con arrossamento della caviglia, presenza di pigmentazione emoderemica e un'anamnesi di vene varicose. Il primo è una donna di 71 anni che per il resto sta bene, il secondo è un uomo di 55 anni con un diabete di tipo 2. Sebbene l'eziologia venosa giustifica più del 70% di tutte le ulcere delle gambe, un clinico esperto sa che la probabilità di base, o pretest, che le ulcere siano di natura venosa, è differente. Per il primo paziente, che ha una manifestazione non complicata, la probabilità pretest di avere un'ulcera venosa è tra circa il 50% e il 70%. Seguendo una semplice valutazione del flusso sanguigno per escludere altre cause, il clinico esperto probabilmente raccomanderà alla paziente di iniziare un trattamento compressivo. In ogni modo, la probabilità pretest di ulcera venosa nel secondo paziente è, probabilmente, considerevolmente inferiore. Le malattie venose causano solamente dal 6% al 9% delle ulcere alle gambe nei pazienti con diabete. Una semplice valutazione, per escludere altre cause, può non convincere il clinico sulla natura venosa dell'ulcera. Il trattamento per le ulcerazioni venose include l'applicazione dei bendaggi fortemente compressivi all'arto compromesso del paziente, ma se il paziente ha un'insufficienza arteriosa il bendaggio può creare un'ischemia alla gamba. Il rischio clinico di una diagnosi mal posta e dell'ischemia può aumentare l'attenzione nell'iniziare il trattamento e la bassa probabilità pretest significa che il clinico può preferire di rinviare il paziente a ulteriori test prima di essere sicuro che la compressione sia salutare. L'esempio illustrato prima, dove non aveva importanza quale fosse il risultato di una valutazione

o di un test, non può indicare ad un clinico se il paziente presenta, oppure no, la malattia. Può solo fornirgli la probabilità che l'abbia o non l'abbia. L'abilità nel discriminare tra le persone che, probabilmente, hanno la malattia e quelle che hanno meno probabilità di averla, è determinata da un esame dell'indice di probabilità (likelihood ratio LR). Rispetto allo screening sulla depressione, Whooley e al. hanno trovato diversi strumenti con risultati simili, ma lo strumento più semplice era lo strumento a due domande. Il test di riferimento indicava che 97 pazienti avevano la depressione. Lo strumento di accertamento del caso a due domande identificava correttamente, come probabilmente depressi, 94 di questi pazienti (94/97 o 0,97). Nello stesso modo, lo strumento classificava in modo scorretto, come probabilmente depressi, anche 189 pazienti, su 439 (189/439 o 0,43) che il test di riferimento escludeva essere depressi (vedi tabella 1).

Risultato del test	DEPRESSIONE		Indice di probabilità
	Presente	Assente	
Presente	0,97	0,43	2,25
Assente	0,03	0,57	0,05
Totale	1,00	1,00	

Tabella 1 risultati dello strumento a due quesiti come strumento di accertamento del caso per la depressione

Il rapporto tra queste due probabilità è il LR. Considerando il LR questo è la percentuale o la proporzione di pazienti che il test identifica, correttamente o scorrettamente, come malati, cioè non considera il reale numero dei pazienti. Così, il rapporto dei risultati veramente positivi (per esempio quelli che lo strumento identifica correttamente come depressi), con i risultati falsi positivi (quelli che lo strumento identifica scorrettamente come depressi), è 0,97/0,43 o 2,25. Questo è l'indice di probabilità che un risultato positivo del test (+LR) sia corretto. Dal +LR 2,25 possiamo dedurre che un risultato positivo dello strumento a due domande aveva solo circa 2 probabilità in più di essere un risultato veramente positivo, piuttosto che essere un falso positivo. Se questo strumento fosse stato usato per la diagnosi della depressione, i clinici sarebbero stati in errore abbastanza frequentemente. Chiaramente lo strumento di accertamento del caso a due domande non è molto efficace per la diagnosi del paziente depresso.

Come può essere calcolato il +LR, può essere calcolata anche la probabilità che lo strumento sia in errore quando fornisce un risultato negativo. Lo strumento a due domande perde 3 dei 97 pazienti depressi (3/97 o 0,03), ma identifica correttamente 250 pazienti, come probabilmente non depressi, su 439 pazienti (250/439 o 0,57) nei quali la depressione è assente. L'indice dei risultati falsi negativi (per esempio quelli che lo strumento

identifica erroneamente come non depressi), rispetto ai risultati veramente negativi (quelli che lo strumento identifica correttamente per non essere depressi), è 0,03/0,57 o 0,05. Questo è l'indice di probabilità che un risultato negativo del test (-LR) sia sbagliato. Dal -LR 0,05 possiamo dedurre che pochissimi pazienti sono probabilmente depressi quando lo strumento di accertamento del caso fornisce un risultato negativo.

L'utilità dei LR si manifesta quando osserviamo la loro capacità di far slittare un paziente da una probabilità pre test a una post test, e facendo ciò, aiutare a ridurre l'incertezza clinica associata agli screening, all'accertamento del caso o alla diagnosi. Nella tabella 2 si può trovare una guida sommaria sulla grandezza dei LR e sul loro effetto sulla probabilità post test.

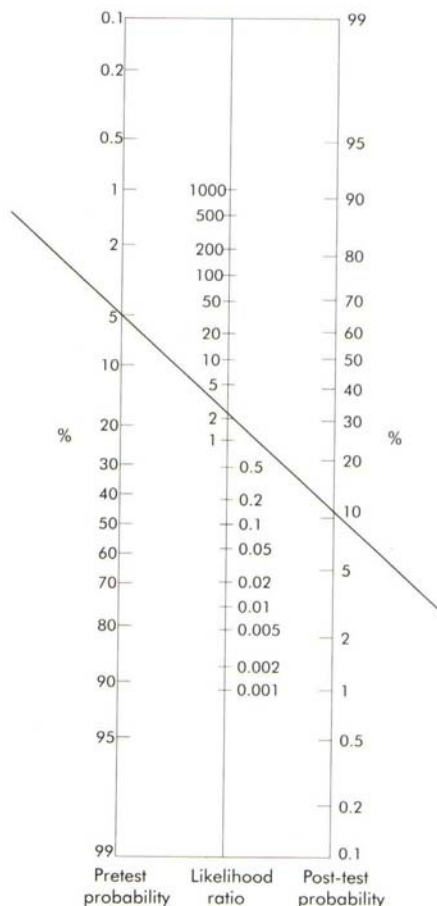
Valore del LR	Utilità
+LR>10 o -LR<0,1	Genera grandi cambiamenti tra le probabilità pre e post test
+LR 5-10 o -LR 0,1-0,2	Genera un moderato spostamento nella probabilità pre test rispetto a quella post test
+LR2-5 o -LR 0,5-0,2	Genera piccoli cambiamenti nella probabilità pre test rispetto a quella post test
+LR 1-2 o -LR 0,5-1	Genera piccoli o nessun cambiamento nella probabilità pre test rispetto a quella post test

Tabella 2 valore dell'indice di probabilità (LR) e associata utilità dei cambiamenti di probabilità

La sfida nel trovare le modificazioni, dopo l'esecuzione di un test, della probabilità che un paziente abbia una malattia è semplificata da un semplice normogramma.

Tracciando una linea dritta tra la probabilità pre test (colonna a sinistra) e il LR (colonna centrale), la probabilità post test può essere determinata dal punto nel quale la linea interseca la colonna di destra. Una probabilità pretest potrebbe semplicemente essere la prevalenza della depressione nel territorio, che è stata stimata essere 5% nella popolazione adulta in Gran Bretagna. Se il paziente nel nostro scenario risponde affermativamente ad entrambe le domande possiamo estendere la linea da una probabilità pretest del 5% fino ad approssimativamente 2 (+LR2,25) per ottenere una probabilità post-test, che il nostro paziente sia veramente depresso, di un po' meno del 10%. Comunque, se il nostro paziente risponde no ad entrambe le domande, possiamo estendere la linea dal 5% attraverso 0,05 (-LR) per ottenere una probabilità post test di essere in errore, se consideriamo il nostro paziente non depresso, di circa lo 0,03%. Quindi,

possiamo essere tranquilli che se un paziente risponde no alle 2 domande è molto improbabile sia depresso. D'altro canto, una probabilità post test di approssimativamente il 10% potrebbe non essere sufficientemente alta, tanto da farci considerare di sottoporlo ad ulteriori accertamenti, a meno che non ci sia un'altra probabile spiegazione dei sintomi del paziente. Comunque, lo strumento può essere utile per determinare se siano auspicabili ulteriori accertamenti in un contesto dove la probabilità pretest è più alta.



Nomogram for determining posttest probabilities.¹¹
 Reproduced with permission from Fagan TJ. *N Engl J Med* 1975;293:257.

Whooley et al. forniscono il LR per ciascuno strumento di accertamento del caso. Gli studi più vecchi spesso non forniscono il LR, ma, al suo posto, forniscono la sensibilità e la specificità del test. La sensibilità di un test è la proporzione di pazienti, che presentano la malattia in esame, che hanno un risultato positivo del test, mentre la specificità è la proporzione dei pazienti, senza la malattia in esame, che hanno un risultato negativo del test. Gli LR sono ottenibili in modo semplice se si conoscono la sensibilità e la specificità di un test. La sensibilità e la specificità dello strumento di accertamento del caso a due domande sono rispettivamente 0,97 e 0,57. Un +LR si ottiene dalla formula seguente:

$$\text{sensibilità} \quad 0.97 \quad 0.97$$

$$\frac{\text{-----}}{1 - \text{specificità}} = \frac{\text{-----}}{1 - 0.57} = \frac{\text{-----}}{0.43} = 2.25$$

Nello stesso modo possiamo ottenere -LR usando una formula poco differente:

$$\frac{1 - \text{sensibilità}}{\text{specificità}} = \frac{1 - 0.97}{0.57} = \frac{0.03}{0.57} = 0.05$$

Qualche volta la sensibilità e la specificità sono presentate come percentuali (per esempio 97% e 57%). Le stesse formule possono essere usate sostituendo 100 a 1 quando si sottrae. Per ulteriori spiegazioni su come la sensibilità e la specificità siano calcolate, guardate Sackett e al. o qualsiasi test sull'epidemiologia clinica.

POSSO APPLICARE QUESTO TEST AL MIO PAZIENTE?

Abbiamo stabilito che lo studio di Whooley e al. è probabilmente valido e decidiamo che i risultati indicano che lo strumento può essere utile per identificare i pazienti che possono beneficiare di essere sottoposti ad ulteriori accertamenti, quando il paziente risponde positivamente alle domande, ed è utile per escludere la depressione come una possibilità, quando il paziente risponde negativamente alle domande. Il prossimo passo è determinare se può essere utilizzato per il vostro attuale paziente o gruppo di pazienti. Potrà aiutarvi in questa decisione rispondere a tre domande.

IL TEST È DISPONIBILE, ACCESSIBILE, ACCURATO E PRECISO PER IL VOSTRO CONTESTO?

Ovviamente, se un test non è disponibile, o i costi sono simili a quelli per alternative più accurate e utilizzabili, è improbabile che venga usato. Ugualmente, abbiamo bisogno di assicurarci che il test manterrà la sua accuratezza nell'ambiente clinico in cui lavoriamo. Gli LR possono essere fissi, ma derivano da selezioni di pazienti e quindi possono non essere accurati per i pazienti che sono stati selezionati in modo diverso. Con una prima domanda sulla validità, abbiamo bisogno di essere rassicurati che lo strumento sia stato testato in pazienti che sono in condizioni medie, moderate e severe così come in quelli senza la malattia. Abbiamo bisogno di essere rassicurati sulla somiglianza della popolazione studiata con quella del nostro ambiente. Non è comune trovare un lavoro che descriva esattamente una popolazione di pazienti come i nostri, così bisogna che esaminiamo la demografia dei partecipanti allo studio, per decidere se sono così diversi dai nostri da escluderne l'uso. Un'altra questione sull'accuratezza di un test è che molti strumenti sono descritti come se avessero solo un

+LR e un -LR, sebbene un test può comportarsi diversamente in base alla severità della malattia. Degli LR più alti si ritrovano in condizioni floride e più bassi in presentazioni precoci della malattia. Alcuni test fanno questa distinzione descrivendo i LR di differenti presentazioni della malattia, ma questo sarebbe inusuale per uno strumento di screening o di accertamento del caso.

POSSIAMO COSTRUIRE STIME CLINICAMENTE SENSIBILI DELLE PROBABILITÀ' PRETEST DEI PAZIENTI?

La probabilità pretest è la probabilità che un paziente attuale abbia un disordine particolare. Sackett e al. identificano differenti fonti per stimare la probabilità pretest: esperienza clinica, prevalenza statistica, database pratici, studi specificatamente focalizzati sulla determinazione delle probabilità pretest e lo stesso studio originale. L'esperienza clinica produrrà essenzialmente una "congettura" e diversi falsi euristici possono influenzare una stima. Comunque, in assenza di altre fonti, anche questo metodo può essere utile. La prevalenza statistica può essere determinata dai dati di morbilità regionali o nazionali, o da studi che esaminano la prevalenza di una malattia, ma queste stime sono buone solo se lo sono le fonti dei dati o il contesto degli studi di prevalenza. I database che derivano dagli inserimenti volontari possono avere dati poco accurati. Se lo studio della prevalenza è attuato in un contesto di assistenza per acuti, i risultati possono essere svianti se applicati in un contesto di assistenza primaria. Anche i database pratici, se locali, regionali o nazionali, sono buoni se lo sono la loro fonte dei dati. Gli studi che esaminano le probabilità pretest sono inferiori come numero e difficili da reperire nei database. Infine, la prevalenza della malattia nello studio deve essere valutata criticamente per poter essere usata.

IL RISULTATO DELLA PROBABILITA' POST-TEST PUÒ INFICIARE LA GESTIONE DEL PAZIENTE?

La preoccupazione maggiore in questo caso è se i risultati porterebbero il paziente al di là di un limite che fa fermare ulteriori accertamenti per la malattia sospettata. Questo dovrebbe accadere quando una malattia è stata esclusa, quando è stato fatto un rinvio ad ulteriori accertamenti o quando è stato iniziato il trattamento. Per esempio, se la probabilità pretest della depressione è del 5% e la risposta del paziente alle due domande dello strumento di accertamento del caso è negativa, la probabilità post-test dovrebbe essere così bassa che la depressione potrebbe essere scartata come spiegazione dei sintomi del paziente. Comunque, se la risposta del paziente è positiva (e ricordando che la probabilità post-test è leggermente

superiore al 10%), sarebbe ancora troppo bassa per spostare il paziente su un trattamento limite, o forse anche per rimandarlo a futuri accertamenti. D'altro lato, se la probabilità pretest è più alta, forse perché vi è un'alta prevalenza di depressione nelle persone con diabete, il rinvio a ulteriori accertamenti potrebbe essere giustificato.

RISOLUZIONE DELLO SCENARIO

Sebbene accettate che lo studio di Whooley e al. abbia una validità ragionevolmente forte, avete delle riserve nell'usare lo strumento di accertamento del caso a due domande con tutti i vostri pazienti. Il campione dello studio è formato principalmente da uomini (97%), 71% sono disoccupati, e il contesto è una clinica d'assistenza urgente, dove i pazienti hanno un alta prevalenza di depressione (18%), invece che una popolazione territoriale, dove la prevalenza è probabilmente molto più bassa. Quindi decidete di continuare a cercare uno strumento di accertamento del caso che sia stato valutato in popolazioni differenti e possa essere maggiormente applicabile ai vostri pazienti. Fino a quando non troverete uno strumento di questo tipo, lo strumento a due domande potrebbe essere utile per questo particolare tipo di paziente. Avete la sensazione che la depressione sia più frequente nelle persone con diabete che nella popolazione in generale e una veloce ricerca sulla letteratura vi permette di trovare una revisione sistematica, sulla prevalenza della depressione nelle persone diabetiche, che conferma la vostra visione. L'indice medio della depressione attuale (all'opposto della storia di depressione) in studi controllati è identificato nel 14%, almeno tre volte quello della popolazione in generale. Decidete di usare questa come la vostra probabilità pretest. Un controllo veloce del normogramma vi mostra che, se il paziente risponde sì ad entrambe le domande, la probabilità post-test di depressione sarà approssimativamente del 27%, che è abbastanza alta per suggerire la necessità di ulteriori indagini. Dato che lo strumento a due domande non è più invasivo di un prelievo di sangue, decidete di discutere con il vostro paziente, la prossima volta che lo andrete a visitare per cambiare la medicazione della sua ferita, la possibilità di una causa clinica per i suoi sintomi. Se sarà disposto, vi risolverete ad utilizzare lo strumento di accertamento del caso a due domande per evidenziare la depressione.

Bibliografia: vedi articolo originale

*ANDREW JULL, RN, MA
Clinical Trials Research Unit,
University of Auckland
Auckland, New Zealand*



Novembre 2002