

Limiti decisionali e algoritmi diagnostici

F. Carle

Centro di Epidemiologia, Biostatistica e Informatica medica, Università Politecnica delle Marche, Ancona

Riassunto

Le indagini di laboratorio rappresentano una fonte preziosa di informazioni per il medico sia nell'ambito dei processi diagnostici che nel monitoraggio dell'evoluzione della malattia nel paziente. L'informazione si genera attraverso un processo di ragionamento logico e scientifico, basato su prove di efficacia, che partendo dal dato di laboratorio, ovvero dal risultato del test, attribuisce a questo ultimo un significato specifico nell'ambito del protocollo diagnostico o prognostico in cui il test è stato richiesto. All'interno del processo vengono applicati alcuni strumenti decisionali come i valori dei test diagnostici che permettono di individuare/escludere la malattia e gli algoritmi che descrivono e valutano le serie di indagini che conducono alla diagnosi; la definizione e la costruzione di questi strumenti può essere complessa.

In questa relazione vengono individuati alcuni punti di riflessione metodologica riguardo la definizione e l'applicazione dei limiti decisionali e la costruzione degli algoritmi diagnostici all'interno del processo logico scientifico che conduce dal dato di laboratorio alla decisione clinica.

Summary

Decision limits and diagnostic algorithms

Laboratory tests are a valuable source of information for the physician both within the diagnostic process and in monitoring disease evolution. The information is generated through an evidence-based process of logical scientific reasoning, that starting from the laboratory data confers on the test result a specific meaning within the diagnostic or prognostic protocol, in the framework of which the test has been applied. Some decisional tools, like the values of the diagnostic tests that identify or exclude the disease and the algorithms that describe and analyse the series of actions that lead to the diagnosis, are applied in this logical scientific process. Defining these tools can be complex. This work reports some methodological reflections about the definition and application of the decision limits and the construction of the diagnostic algorithms within the logical scientific process that from the laboratory data lead to the clinical decision.

Introduzione

Le indagini di laboratorio rappresentano una fonte preziosa di informazioni per il medico sia nell'ambito dei processi diagnostici che nel monitoraggio dell'evoluzione della malattia nel paziente.

L'informazione si genera attraverso un processo di ragionamento logico e scientifico, basato su prove di efficacia, che partendo dal dato di laboratorio, ovvero dal risultato del test, attribuisce a questo ultimo un significato specifico nell'ambito del protocollo diagnostico o prognostico in cui il test è stato richiesto.

Generalmente il processo logico inizia con il con-

fronto del risultato del test con un valore, o un intervallo di valori, che rappresenta il livello decisionale per individuare/escludere la presenza di una determinata condizione nel paziente; in seguito il risultato viene interpretato considerando le altre informazioni a disposizione relative alle caratteristiche del paziente, ai segni e sintomi e ai risultati di altre indagini di laboratorio, strumentali, cliniche. Il valore di glicemia (es. 140 mg/dL), misurato in un campione ematico prelevato a digiuno, assume un significato diagnostico per la presenza del diabete mellito se confrontato con il livello decisionale pari a 126 mg/dL definito dalle linee guida

internazionali¹, ma l'informazione deve essere completata considerando se il test è stato eseguito per la prima volta o se è un test di conferma di un risultato precedente, se il test è stato eseguito in un paziente in cui era già nota la diagnosi di diabete o se rappresenta il primo dato di laboratorio disponibile, se, in questo ultimo caso, sono stati indagati altri fattori di rischio per la malattia diabetica, ecc.

Come noto l'informazione proveniente da un singolo test diagnostico può essere errata e/o incompleta; la sempre maggior disponibilità di test diversi per la stessa patologia conduce alla necessità di disporre di modelli di analisi complessa per potere combinare le informazioni da test multipli e disporre quindi di un miglior strumento diagnostico rispetto a ogni singolo test. Un classico esempio è rappresentato dai biomarcatori tumorali; nessun singolo biomarcatore è sufficientemente sensibile e specifico per la diagnosi precoce del carcinoma dell'ovaio: la ricerca sta cercando di individuare un panel di biomarcatori sierologici adeguato²

L'obiettivo di questa relazione è quello di individuare alcuni punti di riflessione metodologica riguardo la definizione e l'applicazione dei limiti decisionali e degli algoritmi diagnostici all'interno del processo logico scientifico che conduce dal dato di laboratorio alla decisione clinica.

Limiti decisionali o livelli decisionali?

Il primo punto di riflessione riguarda la definizione del termine "limiti decisionali". Questo termine è stato utilizzato per la prima volta dal College of American Pathologists nel 1985³ per definire un intervallo di risultati dei test di laboratorio clinicamente accettabili; il concetto di base espresso dagli autori è che i limiti decisionali comprendano i requisiti clinicamente utili di un test diagnostico in modo che inserendo tali limiti in una carta di controllo di qualità del test stesso, sia possibile individuare il significato clinico oltre che gli errori statistici dei risultati ottenuti. Negli anni successivi tale termine è stato usato in letteratura secondo diverse accezioni e spesso come sinonimo di "livelli decisionali". In realtà tale termine si riferisce a un concetto ben chiaro non solo per la medicina di laboratorio ma anche per la pratica clinica: il livello decisionale è un valore o una concentrazione di una sostanza che discrimina i soggetti in base alla presenza o assenza di una condizione patologica; è importante sottolineare che tale suddivisione è funzione della sensibilità, specificità (validità) e precisione del test.

Ciò conduce a un secondo punto di riflessione. E' possibile individuare due livelli di validità del test: la prima riguarda la capacità del test di individuare la quantità o la condizione che deve misurare (ad esempio la concentrazione reale di troponina cardiaca in ng/ml o la presenza di una mutazione di un oncogene specifico), la seconda la capacità del test di individuare i soggetti con la condizione patologica di cui il valore del

test è un indicatore (ad esempio i soggetti con infarto del miocardio o i soggetti affetti da tumore della mammella). E' intuitivo che i due livelli di validità non sono indipendenti, tuttavia è possibile che un test sia caratterizzato da un'elevata validità nel misurare ciò che deve misurare ma che sia un indicatore di malattia poco valido. Considerando la definizione del College of American Pathologists, il secondo livello di validità andrebbe applicato alla valutazione dei limiti decisionali, mentre il primo rimane caratteristico dei livelli decisionali.

Nel tentativo di chiarire la differenza tra i termini, Westgard⁴ applica il concetto di limiti decisionali alle carte di controllo di qualità di Levey-Jennings e dimostra che questi limiti corrispondono comunque sempre a una regola decisionale basata su un calcolo statistico (per esempio la media più o meno la deviazione standard moltiplicata per due) e quindi non assicurano la qualità clinica necessaria per l'applicazione del test diagnostico ma semplicemente rappresentano dei valori di riferimento (livelli decisionali) più ampi. Egli definisce i limiti decisionali come un mito moderno privo di giustificazioni scientifiche.

In un recente articolo⁵, la definizione di limite decisionale viene discussa in termini di medicina basata sulle prove di efficacia, prendendo come esempio il limite decisionale per l'applicazione del dosaggio delle troponine cardiache (cTnI; cTnT) nella diagnosi dell'infarto del miocardio definito e raccomandato dalle linee guida internazionali⁶.

Queste ultime indicano come limite decisionale il 99esimo percentile della distribuzione dei valori di troponina in una popolazione sana di riferimento richiedendo a tale livello una precisione adeguata identificata con un valore del coefficiente di variazione inferiore o uguale al 10%, allo scopo di minimizzare la quota di risultati falsamente positivi. Questo esempio conduce ad alcune considerazioni di ordine generale.

La prima riguarda la precisione del test; è possibile che i metodi di dosaggio del biomarcatore non siano in grado di assicurare la precisione richiesta al livello del 99esimo percentile e che quindi in realtà venga utilizzato come limite decisionale il valore del test caratterizzato dal coefficiente di variazione richiesto; il limite decisionale corrisponde quindi a un valore di percentile che può variare oltre che tra laboratori diversi anche all'interno dello stesso laboratorio in sedute diverse e in funzione del tipo commerciale del metodo di dosaggio⁷. E' evidente che in questa situazione l'utilità nella pratica clinica del limite decisionale è scarsa in quanto si dispone di un criterio non riproducibile.

E' necessario inoltre ricordare che l'utilizzo del coefficiente di variazione presuppone che la distribuzione dei risultati del test sia simmetrica rispetto alla media (assunto di normalità); come noto, tale assunto dipende dal numero di osservazioni e quindi può non essere verificato se il numero delle ripetizioni del test utilizzato per determinarne la precisione è limitato. Il calcolo della numerosità minima necessaria, che può essere ef-

fettuato con opportuni metodi di analisi statistica, aumenta la solidità del valore del coefficiente di variazione come misura di precisione del test diagnostico.

La seconda considerazione riguarda l'identificazione della popolazione di riferimento che è diversa per ciascun laboratorio e che deve essere definita da questo sulla base delle caratteristiche che potrebbero influenzare i valori del test, come il genere, l'età, l'etnia, le abitudini di vita, ecc.

La terza considerazione riguarda le evidenze scientifiche su cui si basa la scelta del limite decisionale; nell'esempio utilizzato la scelta deriva dalla discussione di un comitato di esperti (all'interno di una consensus conference) e non dai risultati di studi clinici e gli autori stessi suggeriscono la necessità che la scelta effettuata sia confermata attraverso studi clinici di valutazione del rischio di infarto associato a determinati valori di troponina^{8,9}.

Dalle considerazioni sopra esposte e dall'utilizzo del termine in letteratura è possibile tentare di definire i limiti decisionali in base alla loro specifica funzione che sembra essere quella di predire una determinata condizione patologica piuttosto che quella di identificare i soggetti malati.

In altri termini, dopo aver accertato la validità del test diagnostico, o della combinazione di test diagnostici, identificando i valori di riferimento in grado di discriminare i soggetti malati dai soggetti non malati, l'interesse clinico si rivolge verso la possibilità di stimare la probabilità (rischio) di una certa condizione patologica in presenza di un determinato valore di un test diagnostico, ovvero verso la capacità predittiva di questo ultimo. E' quindi necessario individuare i valori del test caratterizzati dalla massima probabilità di predire la malattia, cioè quei valori che consentono di prendere una decisione riguardo la diagnosi, ovvero occorre identificare i limiti decisionali del test diagnostico.

Come noto il valore predittivo dei risultati di un test è funzione della sensibilità e specificità del test stesso ma anche delle informazioni a priori riguardo le caratteristiche del paziente sottoposto al test. Se queste informazioni vengono tradotte in probabilità stimate nella popolazione da cui proviene il singolo paziente, è possibile costruire un algoritmo decisionale che partendo dai risultati del test diagnostico conduce alla diagnosi della malattia e prosegue fino all'individuazione del percorso terapeutico più vantaggioso per il paziente sia in termini di interruzione del processo patologico che di qualità della vita e di costi.

Dai limiti decisionali agli algoritmi diagnostici

Un algoritmo diagnostico è costituito da una serie di indagini sequenziali intese a risolvere uno specifico problema. In letteratura questo termine viene applicato a due diverse tipologie di problemi. Un algoritmo diagnostico può essere costruito per valutare la validità di uno o più test diagnostici e in questo caso il punto di

partenza è rappresentato dal reclutamento di soggetti per i quali è nota la presenza o assenza della malattia d'interesse individuata con un metodo diagnostico di riferimento (gold standard).

Se invece si vuole costruire uno strumento che permetta di attribuire, sulla base di una serie di informazioni, un soggetto a una certa categoria diagnostica, l'algoritmo può costituire la prima parte di un albero decisionale cioè di un diagramma di flusso che partendo dal quesito diagnostico specifica i percorsi delle diverse azioni diagnostiche e terapeutiche e i loro possibili esiti e permette di stimare l'utilità per il paziente di ogni percorso decisionale alternativo e quindi di scegliere la strategia migliore¹⁰.

E' evidente che la valutazione della validità dei test diagnostici che forniscono le informazioni utili alla diagnosi finale rappresenta la prima fase della costruzione dell'albero decisionale.

La crescente disponibilità di diversi test diagnostici (di laboratorio, strumentali, clinici) per la stessa malattia e la consapevolezza che ogni test può dare risultati errati (falsi positivi e falsi negativi) conduce molto spesso a costruire algoritmi diagnostici che iniziano con la combinazione dei risultati di test multipli allo scopo di migliorare la capacità diagnostica rispetto all'applicazione di un singolo test.

La combinazione dei test viene effettuata utilizzando gli operatori logici Booleani "E" e "O" che, nel caso più semplice costituito dalla combinazione di due test binari (2 risultati possibili: positivo e negativo) per la rilevazione della presenza o assenza di una malattia, identificano 4 possibili configurazioni dei risultati e due criteri decisionali riferiti all'applicazione dei test in serie o in parallelo. Il primo criterio decisionale classifica un soggetto come malato se almeno uno dei due test è positivo (applicazione dei test in parallelo, TP, operatore booleano "O"), mentre con il secondo criterio, la malattia viene identificata se entrambi i test sono positivi (applicazione dei test in serie, TS, operatore booleano "E").

E' possibile dimostrare che le due combinazioni dei test sono caratterizzate da una diversa validità che conduce ad alcune considerazioni importanti per la pratica clinica.

Il criterio TP aumenta la sensibilità e riduce la specificità rispetto ai singoli test, anche se la quota dei falsi positivi della combinazione rimane sempre inferiore alla somma delle rispettive quote dei singoli test; il criterio TS aumenta la specificità assicurando un valore accettabile di sensibilità anche se inferiore alla sensibilità dei singoli test. Ciò implica che il primo criterio decisionale sarà applicato quando si hanno a disposizione test altamente specifici ma scarsamente sensibili, mentre il criterio TS verrà applicato avendo a disposizione test poco specifici. Nell'applicazione in serie la sequenza dei test non influenza la validità della combinazione ma applicando per primo il test più specifico (cioè quello con la quota minore di falsi positivi) si riduce il

numero di soggetti su cui deve essere applicato il secondo test (cioè tutti i soggetti positivi al primo test)¹¹.

Quando i test da applicare sono più di due e/o i risultati dei test sono espressi da una scala continua (per es. i test per i biomarcatori tumorali o miocardici), la metodologia per identificare il criterio decisionale migliore diventa via via sempre più complessa e non è possibile considerare tutti i possibili criteri decisionali allo scopo di trovare quello più adatto, in analogia a quanto descritto per la combinazione di due test dicotomici. La trattazione dettagliata delle soluzioni proposte, di cui si riportano i riferimenti bibliografici^{12,13}, va oltre gli scopi di questa relazione.

Qualunque sia il modello di analisi utilizzato per identificare la combinazione dei test più valida è importante sottolineare due aspetti importanti. Come noto, sensibilità e specificità sono due misure di validità relative, cioè variano in funzione del metodo di riferimento utilizzato come gold standard e applicato in modo indipendente rispetto al test in studio; nel caso dei test multipli è indispensabile che tale indipendenza sia chiaramente verificata per tutti i test della combinazione.

Un secondo aspetto importante riguarda il fatto che i valori calcolati di sensibilità e specificità dei singoli test e delle loro combinazioni rappresentano in realtà delle stime campionarie dei valori veri delle misure di validità di cui è indispensabile valutare la precisione che è tanto migliore quanto più grande è il numero di soggetti che costituiscono il campione studiato. Nella valutazione delle combinazioni è dunque necessario reclutare un numero adeguato di soggetti in funzione oltre che del livello di precisione desiderato anche del numero di test e del tipo di scala con cui sono espressi i risultati (dicotomica, ordinale o continua); in letteratura sono disponibili opportuni metodi per il calcolo della dimensione campionaria necessaria¹⁴.

Algoritmi diagnostici e alberi decisionali

Dopo aver valutato la validità delle combinazioni diagnostiche si prosegue nella costruzione dell'algoritmo stimando la capacità predittiva di malattia dei risultati delle combinazioni stesse; infatti l'obiettivo finale dell'algoritmo diagnostico è quello di valutare quale probabilità ha un individuo di essere assegnato a una certa categoria diagnostica essendo noti i risultati dei test diagnostici a cui è stato sottoposto.

Questa probabilità, definita come probabilità a posteriori di malattia o valore predittivo di un risultato del test, viene stimata attraverso l'applicazione di modelli di analisi bayesiana¹⁵ che permettono di effettuare la stima in funzione dei valori di sensibilità e specificità dei test diagnostici e della probabilità di malattia che ha il soggetto prima di sottoporsi alle indagini diagnostiche previste dall'algoritmo (probabilità a priori).

La probabilità a priori può essere definita come la prevalenza della malattia nella popolazione da cui proviene il paziente, ovvero come il rischio di malattia per gli individui che presentano quelle caratteristiche che

potrebbero modificare la probabilità di malattia del paziente come l'età, il genere, le condizioni e le abitudini di vita, ecc. La stima di questa probabilità viene effettuata attraverso studi epidemiologici condotti su popolazioni diverse e influenza in modo importante il valore predittivo dei test diagnostici.

In presenza di bassi valori di prevalenza, infatti, anche una combinazione di test caratterizzata da un'elevata sensibilità e specificità, può avere una scarsa capacità predittiva di malattia¹⁶. Dal momento che spesso i valori di prevalenza di malattia riferiti a una determinata popolazione non sono disponibili, è utile effettuare un'analisi della sensibilità delle stime della probabilità a posteriori ottenute, valutando come le variazioni della probabilità a priori, compatibili con il problema diagnostico d'interesse, possano influenzare il valore predittivo¹⁷.

L'introduzione dell'algoritmo diagnostico in un albero decisionale consente di rappresentare e quantificare l'interazione spesso complessa tra le decisioni prese dal clinico e lo svolgersi degli eventi, quali i risultati dei test diagnostici e/o le variazioni nelle condizioni del paziente.

L'albero decisionale è lo strumento fondamentale dell'analisi delle decisioni; questa ultima è un processo che consente di scindere la struttura logico-temporale della soluzione di un problema nelle sue componenti, controllando le evidenze e le incertezze che caratterizzano i diversi percorsi che dalla individuazione del problema stesso conducono alle possibili soluzioni¹⁰.

All'algoritmo diagnostico, che individua i rami dell'albero decisionale caratterizzati da una probabilità (per es. la probabilità di risultare positivo o negativo al test diagnostico), vengono aggiunti i rami decisionali che descrivono i percorsi delle diverse azioni (per esempio effettuare o non effettuare un test) e i possibili esiti di queste ultime.

Ciascun esito viene poi quantificato in termini di utilità, definita come una misura della forza della preferenza del paziente per un particolare esito; la quantificazione avviene attraverso l'applicazione di modelli complessi di analisi statistica multipla che considerano la preferenza del paziente e la durata e la qualità della vita.

La combinazione dell'utilità dell'esito e delle probabilità che caratterizzano il percorso che conduce all'esito stesso permette di stimare l'utilità attesa di ogni percorso decisionale alternativo e quindi di scegliere la strategia migliore per il paziente.

L'albero può poi essere ancora completato stimando il costo per ogni percorso decisionale, che può essere importante nella scelta tra due strategie di utilità paragonabile^{10,18}.

Conclusioni

Il processo diagnostico è un processo di ragionamento logico-scientifico complesso in cui il medico definisce, rileva e analizza opportune informazioni ri-

guardo il paziente, la popolazione da cui egli proviene e le possibili alternative diagnostiche.

L'affidabilità di queste informazioni è indispensabile perché il medico possa individuare la giusta diagnosi.

Per quanto riguarda le informazioni fornite dai dati di laboratorio, una quota della affidabilità delle stesse, dipende dai metodi utilizzati nella fase post-analitica per interpretare i risultati dei test diagnostici. L'interpretazione deve essere effettuata considerando tutte le altre informazioni presenti all'interno del processo diagnostico e utilizzando strumenti metodologici adeguati che consentano di analizzare le relazioni tra le informazioni e di stimare le probabilità di errore delle conclusioni raggiunte.

Affinché ciò si verifichi è necessario che le diverse competenze che hanno prodotto le informazioni utili, cioè clinici, medici di laboratorio, patologi e radiologi ma anche epidemiologi e biostatistici, siano a conoscenza delle diverse componenti del processo diagnostico in modo da poter collaborare attivamente nella definizione dello stesso.

Bibliografia

1. The Expert Committee on the diagnosis and classification of diabetes mellitus. Report of the Expert Committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* 2003; 26:S5-S20.
2. Crump C, McIntosh MW, Urban N, Anderson G, Karlan BY. Ovarian cancer tumor marker behavior in asymptomatic healthy women: implications for screening. *Cancer Epidemiol Biomarkers Prev* 2000; 9:1107-11.
3. Howanitz PJ, Kafka MT, Steindel SJ. Quality control run acceptance and rejection using fixed and medically useful limits for QAS Today. *Clin Chem* 1985; 31:1016A.
4. Westgard QC. Essay: the myth of medical decision limits. Disponibile sul sito <http://www.westgard.com/essay8.htm> (ultima consultazione 3 aprile 2006).
5. Morrow DA. Evidence-based decision limits for cardiac troponin: low-level elevation and prognosis. *Am Heart J* 2004; 148:739-42.
6. The Joint European Society of Cardiology/American College of Cardiology Committee. Myocardial infarction redefined - a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J Am Coll Cardiol* 2000; 36:959-69.
7. Galli GA, Gambetta C, Caputo M, Cappelletti P. Indagine sull'utilizzo dei marcatori miocardici in Italia. *RIMeL/IJLaM* 2005; 1:46-55.
8. Wu A, Apple F, Gibler W. National Academy of Clinical Biochemistry Standards of Laboratory Practice: recommendations for the use of cardiac markers in coronary artery diseases. *Clin Chem* 1999; 45:1104-21.
9. Liberati A (a cura di). *La medicina delle prove di efficacia*. Roma: Il Pensiero Scientifico Editore; 1997.
10. Weinstein MC, Fineberg HV. *L'analisi della decisione in medicina clinica*. Milano: Franco Angeli Editore; 1984.
11. Pepe MS. Combining multiple test results. In: *The statistical evaluation of medical tests for classification and prediction*. New York: Oxford University Press; 2003.
12. Pepe MS, Thompson ML. Combining diagnostic test results to increase accuracy. *Biostatistics* 2000; 1:123-40.
13. McIntosh M., Pepe MS. Combining several screening test: optimality of the risk score. *Biometrics* 2002; 58:657-64.
14. Obuchowski NA. Sample size calculations in studies of test accuracy. *Statistics methods in medical research* 1998; 7:371-92.
15. Berry DA, Stangl DK (a cura di). *Bayesian Biostatistics*. New York: Marcel Dekker Inc.; 1996.
16. Harold CS. Probability theory in the use of diagnostic tests. An introduction to critical study of the literature. *Ann. Int. Med* 1986; 104:60-6.
17. Habbema JDF, Boussuyt PMM, Dippel DWJ. Analysing clinical decision analyses. *Statistics in medicine* 1990; 9: 1229-42.
18. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on medical decision analysis: part 1-5. *Medical Decision Making* 1997; 17:123-59.