

EBM: il 2003 è finito

di Giuseppe Giocoli

GdS Amcli sulla EBLM

Paracadute e patacche

“La percezione dell'utilità del paracadute si basa su un'evidenza aneddotica,” esordisce un articolo sul numero di Natale del *British Medical Journal* **(1)** e, “come per molti altri strumenti di prevenzione, la sua efficacia non è stata mai dimostrata con prove rigorose ...”. A scrivere sono i ricercatori Smith e Pell che lanciano un ironico appello - **A call to (broken) arms** - per uno studio comparativo dall'ineccepibile disegno: un braccio sperimentale (discesa con paracadute) e uno di controllo (discesa senza paracadute). “Il reclutamento non sarà facile,” ammettono. “Ma ci rassicura il pensare che il popolo dell' *evidence-based medicine* non esiterà a offrirsi volontario ...”. In realtà l'*humour* riveste un serio messaggio: non è il fondamentalismo scientifico che deve guidare la ricerca delle prove, ma la consapevolezza che l'efficacia di certi interventi, sia terapeutici che preventivi o diagnostici, non è sempre così lampante come per l'uso del paracadute. Uno dei problemi connessi è smascherare chi tinge d'oro il ciarpame sanitario; lo fanno, sullo stesso numero del *British*, Dave Sackett e Andy Oxman **(2)** che, associati nella truffaldina **BALDRACCA srl (HARLOT plc)**, offrono sofisticati sistemi per dimostrare l'efficacia di farmaci patacca o screening inutili, “eludendo la verità senza dire vere bugie”. *It's the money, dummy* (sono soldi, stupido).

L'alba della diagnostica

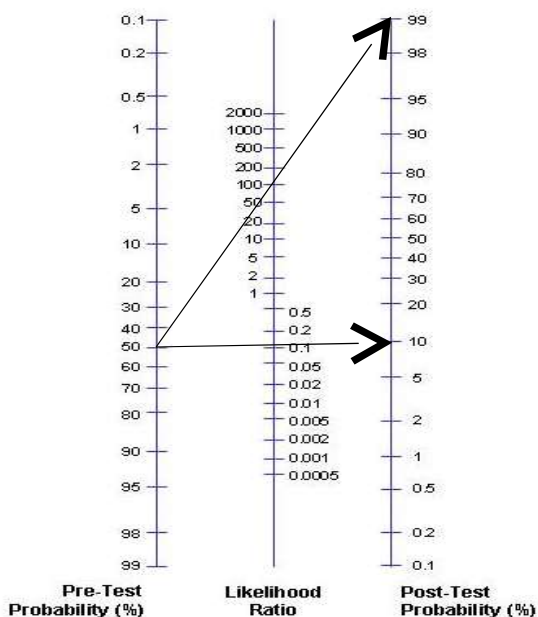
Ma rientriamo nell'accademia con un altro articolo pubblicato in novembre **(3)**. “Quante patologie può ravvisare un generalista?” chiedono Chris Del Mar e Paul Glasziou, rispettivamente dirigenti dei *Centri per la medicina di base* alle università di Queensland e di Oxford. L'occasione per la domanda è la descrizione, in un altro articolo, di un sensibile e semplicissimo test per individuare - nell'ambulatorio di base - il paziente con un possibile stato depressivo. “In diagnostica, *più* non vuol dire sempre *meglio*,” osservano. “Talvolta sono sufficienti pochi test efficaci; l'importante è poter separare il grano dal loglio”. Secondo i ricercatori, siamo all'**alba** di una nuova fase dell'**EBM**: ne è prova la recente decisione della *Cochrane Collaboration* di preparare Revisioni sistematiche su **argomenti diagnostici (4)**. “Man mano che le difficoltà saranno superate,” commentano Del Mar e Glasziou, “cominceremo ad elencare in modo chiaro le cose utili e quelle inutili. Saranno di più queste ultime? E' probabile”.

CDAT per la tubercolosi

“*There is a tendency to believe that molecular tests are intrinsically more likely to be correct because they are new,*” mi scriveva mesi fa Francis Drobniowski, autore di una rassegna sulla moderna diagnosi di laboratorio della tubercolosi **(5)**. E sulla rilevanza in questo settore dei test commerciali ad amplificazione diretta (CDAT) è recentissima una *Minireview* di Claudio Piersimoni e Claudio Scarparo **(6)** che, pur senza menzionarle in modo esplicito, spiegano l'utilità delle **procedure bayesiane**. Cito sintetizzando: “La specificità dei CDAT è elevata ma non la sensibilità: un CDAT negativo non esclude pertanto la tbc e, in sostanza, le decisioni del curante dipendono dal giudizio clinico...”. Dunque: “è nei pazienti con es. microscopico negativo ma elevato sospetto clinico che i CDAT si dimostrano in special modo convenienti”, mentre: “considerati i costi, il loro uso nei pazienti con probabilità a priori di tbc molto basse o molto alte è uno spreco di risorse”. Mi piace infine citare il passo: “... è necessario valutare sensibilità e specificità dal punto di vista clinico oltre che da quello analitico. Molto spesso gli eccellenti risultati della letteratura si riferiscono alle prestazioni analitiche piuttosto che alla diagnosi clinica”. Era ora: sono passati più di dieci anni da quando gli epidemiologi canadesi ce l'hanno detto **(7)**.

Il valore aggiunto di un CDAT

Il **valore aggiunto (VA)** di un CDAT in un caso di microscopia negativa, che non risolve i dubbi per un paziente in forte sospetto di tbc polmonare, può essere spiegato con un esempio:



Con probabilità a priori di tbc equivalenti all'incertezza (50%), un CDAT con sensibilità 88% e specificità 99.2% (LR+ 110, LR- 0.11) **(8)** genera un valore predittivo del 99% ca. (se il risultato è positivo), ma non riesce ad abbassare al di sotto del 10% le probabilità di malattia (se il risultato è negativo), a causa della modesta sensibilità. Conclusione: il VA è limitato ai risultati positivi.

Ma “da dove spuntano **le probabilità a priori**?” L'interrogativo è più che legittimo e Sackett fornisce diverse risposte nel suo aureo libretto **(9)**. Tra esse una *sensitivity analysis* mediante la quale, righello e nomogramma di Fagan alla mano, potremo esplorare un ampio range di probabilità a priori di malattia: ci accorgeremo presto che, disponendo di un test della potenza descritta, è sufficiente una probabilità pre-test del 10% per catapultare il valore predittivo positivo oltre il 90%, mentre per far scendere le probabilità post-test al di sotto dell'1% (ragionevole soglia di esclusione dell'infezione) **(10)** è indispensabile partire da probabilità a priori inferiori al 5% ...

La validità dei test per le malattie respiratorie

La rilevanza dei test molecolari per la ricerca del micobatterio tubercolare nel lavaggio bronchiale è vista con una diversa angolazione in un'analisi retrospettiva della diagnosi e del monitoraggio delle pneumopatie in un'università britannica **(11)**. “Una pratica clinica fondata sull'evidenza scientifica è un obiettivo primario del governo clinico,” affermano Borrill e collaboratori (tra i quali P. Sestini di Siena). “In tale ottica, bisogna convenire che l'efficacia di qualsiasi terapia dipende da una diagnosi accurata. Tuttavia gli studi di accuratezza diagnostica sono spesso mal eseguiti e affetti da errori metodologici”. Dopo un'estesa ricerca bibliografica, gli pneumologi hanno allora classificato gli studi di validazione dei test da loro preferiti secondo **i criteri dell' Oxford Centre for Evidence-based Medicine (CEBM) (12)**. Ebbene, nella scala delle evidenze scientifiche, solo la metà dei test diagnostici e solo un quinto di quelli utili per il monitoraggio sono classificabili ai livelli più alti. In particolare, se l'affidabilità di un risultato positivo per il micobatterio tubercolare nel lavaggio bronchiale è elevata (**livello 1c**, specificità dimostrata), non è altrettanto per l'esclusione dell'infezione se il risultato è negativo. La sensibilità lascia infatti a desiderare e inoltre la validità degli studi effettuati a proposito scende al **livello 4**, perchè minata da difetti inerenti il disegno statistico (studi caso-controllo) o le modalità relative al confronto con il gold-standard.

I livelli di evidenza ed i gradi di raccomandazione

Nel capitolo dall'omonimo titolo, al sito Internet del CEBM di Oxford, si può leggere: “Cosa possiamo fare quando l'inderogabile necessità di fornire una guida al clinico urta contro la mancanza di un'evidenza sicura? Pur facendo del nostro meglio, non ci resta che sottolineare i limiti dei consigli che siamo in grado di fornire” **(12)**. E' questo lo spirito che ha guidato Borrill e colleghi nella loro ricerca **(11)** a conclusione della quale essi segnalano “la necessità di studi di accuratezza diagnostica di profilo più alto”. Ne va della qualità e della stessa possibilità di eseguire le **Revisioni sistematiche (RS)**, lo strumento principe per un' evidenza di massimo livello **(12)**. Finora, a mia conoscenza e per quanto concerne i metodi di amplificazione per la diagnosi di tbc polmonare con microscopia negativa, l'unica indagine con tali requisiti è quella di Sarmiento e coll **(13)**, che utilizzano idonei strumenti epidemiologico-statistici per le loro valutazioni.

La lunga alba dell'EBM in diagnostica risuona delle iniziative della comunità scientifica – ed è doveroso riconoscere i meriti dei ricercatori anglo-americani e olandesi – per agevolare e migliorare l'attuazione delle RS. L'iniziativa **STARD (14)**, da tempo ampiamente propagandata dal nostro gruppo, intende migliorare la costruzione e l'esposizione degli studi diagnostici primari. Un

nuovo documento, il **QUADAS (15)** riguarda le procedure per eseguire le RS in modo standardizzato. Lo propongono esperti che, mediante la procedura Delphi, hanno raggiunto un consenso per una *checklist* di 14 voci, tra le quali: lo spettro clinico dei pazienti, lo standard di riferimento, le modalità per evitare vari tipi di bias, la gestione dei risultati indeterminati. Un altro documento del **CDRH-FDA** sull'analisi di discrepanza **(16)** riguarda invece le modalità per evitare arbitrarie valutazioni dell'accuratezza di un test, quando si ottengono risultati indeterminati. Sono tre esempi importanti, ma certo non i soli.

RT-PCR per la SARS

Parlando di test per le infezioni respiratorie, è forse possibile ignorare la **SARS**? Ai primi di dicembre 2003 il *Canadian Medical Association Journal* ha pubblicato un articolo sull'interpretazione dei test per la SARS alla luce dell'esperienza di Toronto **(17)**, con un commento a cura di Ron Fouchier e Ab Osterhaus, virologi all'*Erasmus Medical Center* di Rotterdam **(18)**. I ricercatori enumerano le molte difficoltà (tuttora presenti) connesse alla **validazione dei primi test molecolari** - come l'assenza di un gold standard e l'eterogeneità nell'esecuzione - e illustrano le analisi retrospettive che, malgrado questi problemi, sono state eseguite per correlare i dati di laboratorio con i dati clinici ed epidemiologici dell'*outbreak* di Toronto. La principale conclusione è che la presentazione clinica e la storia di un contatto rimangono le pietre miliari della diagnosi precoce di SARS. "Nè la prima, nè la seconda generazione della RT-PCR sono in grado di fornire prove definitive d'infezione da SARS-CoV - nè in caso di risultati negativi nè per singoli risultati positivi," sostengono Fouchier e Osterhaus.

In realtà le cose sono alquanto migliorate con i test di seconda generazione, ma "il problema è che - a paragone di altre infezioni respiratorie - il carico virale nella fase iniziale della malattia è molto basso," spiegano; "urge pertanto **migliorare la sensibilità**". Attualmente si stanno sperimentando su larga scala sequenze geniche alternative, ad es. quelle del gene nucleocapsidico in luogo di quelle del gene della replicasi. "Ma è difficile che ciò sconvolga le attuali strategie in caso di nuove epidemie," concludono i virologi. Ecco una malattia per la quale non disponiamo ancora di test sicuri. Ed è forse opportuno meditare sulla realistica posizione del CEBM di Oxford sulla migliore evidenza disponibile **(12)**.

PET per i tumori mediastinici

Quanto tali problematiche siano attuali in tutte le branche della diagnostica lo si può dedurre da un editoriale **(19)** di *Annals of Internal Medicine*, nel quale è commentata una metanalisi sull'impiego della PET (*Positron Emission Tomography*) per lo *staging* del K polmonare con metastasi mediastiniche. I risultati indicano variazioni della sensibilità e della specificità della PET - peraltro elevate nella media - in funzione della gravità della malattia, il tipico "**effetto spettro**" che il menzionato protocollo QUADAS intende ridurre. "La casistica è notevole e le conclusioni della metanalisi appaiono credibili," osservano gli editorialisti. "Ma permangono dei punti oscuri: uno è quello dei **bias** che affliggono gli studi originali e un altro, forse maggiore, è l'interrogativo su **reali prestazioni ed effetti della PET nella routine** ...". Secondo i ricercatori, iniziano ora ad essere utilizzati in campo diagnostico i trial randomizzati e controllati: "è questo il disegno dello studio che può chiarire **l'impatto clinico** di differenti strategie alternative". Nell'editoriale non manca una descrizione delle principali tappe del lungo viaggio di un test diagnostico verso la validazione e la dimostrazione di efficacia; "... non è prudente dunque introdurre nuovi test in assenza di dati sicuri sugli *outcomes* finali e sulla sola base della sensibilità e della specificità".

Knowledge brokering

Microbiologi, ricordate la sessione sull'appropriatezza al congresso AMCLI di Firenze? G. Ippolito ci ha parlato del trasferimento della conoscenza *from bench to bedside*, un traguardo raggiungibile (secondo l'infettivologo) in una nuova generazione, quando l'approfondimento e la consuetudine avranno fatto superare "l'infarinatura che caratterizza l'attuale livello di conoscenza di tali problemi" **(20)**. Ebbene, i nostri amici d'oltre-oceano non perdono tempo: è del dicembre 2003 un rapporto della *Fondazione canadese per la ricerca in sanità pubblica* (CHSRF) che chiarisce parecchi aspetti dell'argomento **(21)**. Ne cito alcuni passi: "Esiste un processo, il **brokering della conoscenza**, che consiste nell'acquisire e stivare le cognizioni per poi condividerle sotto la direzione di specialisti - i **broker** - in vista di nuovi utilizzi e nuove applicazioni ..." (p.1 del testo cit.). Ma chi sono esattamente questi *broker*? Letteralmente *mediatori*, il CHSRF li definisce "negoziatori che facilitano la comunicazione, l'accesso all'informazione e lo scambio di conoscenze tra i componenti di un sistema finalizzato (*network*)" (p.4).

In altre parole, non è un un *knowledge broker* lo specialista che si prende la briga di aggiornare colleghi e amministratori sulle novità scientifiche, perchè non fa che trasferire conoscenza. "*The broker role is about bringing people together*": lo diviene dunque "se organizza riunioni con i suoi colleghi e i responsabili della politica sanitaria, aiutandoli a sviluppare progetti basati sull'evidenza scientifica".

Evidence-brokering, appunto. Buon lavoro e buon 2004.

Giuseppe Giocoli
GdL EBM – AMCLI
1 gennaio 2004

Bibliografia

- 1.** GCS Smith e JP Pell. Hazardous journey. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ* 2003;327:1459-61 (20 December)
- 2.** DL Sackett, AD Oxman on behalf of HARLOT plc. HARLOT plc: an amalgamation of the world's two oldest professions. *BMJ* 2003;327:1442-5 (20 December)
- 3.** C Del Mar e P Glasziou. How many conditions can a GP screen for? Evidence based information on diagnostic test will help. *BMJ* 2003;327:1117 (15 November)
- 4.** J Deeks et al. Cochrane systematic reviews of diagnostic test accuracy. The Cochrane Collaboration Methods Groups Newsletter 2003; 7: 8 www.cochrane.de/newslet/MGNews_2003.pdf
- 5.** F A Drobniewski et al. Modern laboratory diagnosis of tuberculosis. *Lancet Inf Dis* 2003; 3:1141-7
- 6.** C Piersimoni e C Scarparo. Relevance of Commercial Amplification Methods for Direct Detection of Mycobacterium tuberculosis Complex in Clinical Samples. *J. Clin. Microbiology* Dec. 2003; 41: 5355-65
- 7.** DL Sackett et al. *Clinical Epidemiology. A basic science for clinical medicine* 1991 (pag 63, della 2.a ed). Little Brown & Co.
- 8.** C Piersimoni et al. Performance Assessment of Two Commercial Amplification Assays for Direct Detection of Mycobacterium tuberculosis Complex from Respiratory and Extrapulmonary Specimens. *J. Clin. Microbiology* 2002; 40: 4138-42
- 9.** DL Sackett et al. *Evidence-based Medicine* (1st Ed) 1998, p. 118. Churchill Livingstone
- 10.** Anonimo. Tuberculosis Diagnosis. *Bandolier* 79-7. Sept. 2000 <http://www.ebandolier.com>
- 11.** Z Borrill et al. Retrospective analysis of evidence base for tests used in diagnosis and monitoring of disease in respiratory medicine. *BMJ* 2003;327:1136-8 (15 November)
- 12.** Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001) www.cebm.net/levels_of_evidence.asp
- 13.** OL Sarmiento, et al. Assessment by meta-analysis of PCR for diagnosis of smear-negative pulmonary tuberculosis. *J. Clin. Microbiology* 2003; 41:3233-40.
- 14.** The STARD Group. Towards complete and accurate reporting of studies on diagnostic accuracy: the STARD initiative. <http://www.consort-statement.org/stardstatement.htm>
versione italiana: <http://www.gimbe.org/Link/STARD.zip>
- 15.** P Whiting et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Medical Research Methodology* 2003;3:25
- 16.** CDRH - Center for Devices and Radiological Health
. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests; Draft Guidance for Industry and FDA Reviewers 2003 <http://www.fda.gov/cdrh/osb/guidance/1428.pdf>
- 17.** Tang P, et al. Interpretation of diagnostic laboratory tests for severe acute respiratory syndrome: the Toronto experience. *CMAJ* [early online release 2003 Dec 2]. Available: www.cmaj.ca

- 18.** Ron AM Fouchier, Ab DME Osterhaus. Laboratory tests for SARS: Powerful or peripheral? CMAJ Jan 6, 2004; 170 [early online release 2003 Dec 2]. Available: www.cmaj.ca
- 19.** RA Deyo e JJ Jarvik. New Diagnostic Tests: Breakthrough Approaches or Expensive Add-ons? Editorial. Ann Intern Med. 2003;139:950-1
- 20.** G.Ippolito. *Translational Clinical research e Technology assessment* in Microbiologia. Sess. 1 del XXXII Congr Naz AMCLI. Firenze, ottobre 2003.
- 21.** CHSRF - Canadian Health Services Research Foundation. The Theory and Practice of Knowledge Brokering in Canada's Health System. December 2003 www.chrsf.ca