



**ASSOCIAZIONE
OTORINOLARINGOLOGI
OSPEDALIERI
ITALIANI**
Presidente: **CARLO ANTONIO LEONE**

METODOLOGIA CLINICA APPLICATA IN O.R.L.

**(Evidence Based Medicine, analisi critica della letteratura,
ricerca bibliografica in banche dati elettroniche)**

a cura di

**MARCO PIEMONTE
STEFANO PALMA**

torgraf

QUADERNI MONOGRAFICI DI AGGIORNAMENTO



ASSOCIAZIONE
OTORINOLARINGOLOGI
OSPEDALIERI
ITALIANI

Presidente: CARLO ANTONIO LEONE

METODOLOGIA CLINICA APPLICATA IN O.R.L.

**(Evidence Based Medicine, analisi critica della letteratura,
ricerca bibliografica in banche dati elettroniche)**

a cura di

Marco Piemonte

e

Stefano Palma

QUADERNI MONOGRAFICI DI AGGIORNAMENTO

© Quaderni Monografici di Aggiornamento A.O.O.I.

METODOLOGIA CLINICA APPLICATA IN O.R.L.

a cura di

Marco Piemonte e Stefano Palma

S.O.C. ORL, Az. Ospedaliero-Universitaria
"S. Maria della Misericordia" Udine



La riproduzione di questo volume o di parte di esso e la sua diffusione in qualsiasi forma o con qualsiasi mezzo, elettronico, meccanico, per mezzo di fotocopie, microfilm, registrazioni od altro, sono proibite senza il permesso scritto della A.O.O.I. (Associazione Otorinolaringologi Ospedalieri Italiani).

Realizzazione editoriale e stampa:

TorGraf

S.P. 362 km. 15.300

73013 Galatina (Le)

Telefono 0836.561417

Fax 0836.569901

e-mail: stampa@torgraf.it

PRESENTAZIONE

È con piacere che presento questo volume dei quaderni di Aggiornamento pubblicati dalla Associazione scientifica che ho l'onore di presiedere.

Si tratta di un volume assolutamente attuale che vede per la prima volta una esposizione sistematica e completa sulla metodologia clinica applicata in ORL.

La corretta impostazione di un contributo scientifico necessita di una metodologia clinica con solide basi inerenti la tecnica di raccolta dei dati clinici, l'inquadramento epidemiologico, la discussione e la loro analisi critica: l'attento rispetto di questi elementi facilita il confronto tra scuole di pensiero differenti.

L'importanza dell'argomento non scaturisce soltanto da una compilazione corretta di pubblicazioni utili alla formazione di singoli professionisti, ma si estende anche al fatto che la mole complessiva del prodotto scientifico della specialità finisce per influenzare i comportamenti clinici di una collettività di medici e indirettamente orientare alcune scelte organizzative degli amministratori della sanità.

L'incremento delle tecnologie informatiche e la loro diffusione ha generato una facilitazione ed una maggiore rapidità di scambio di informazioni che sono alla base del progresso scientifico. Questo miglioramento tecnologico deve coniugarsi ad una metodologia clinica adeguata.

Pertanto con queste premesse mi auguro che lo sforzo dei colleghi ed amici Marco Piemonte e Stefano Palma sia accolto col massimo favore da quanti di noi desiderino migliorare le proprie conoscenze e condividere i propri contributi scientifici rendendoli disponibili e fruibili per tutti ed in particolare per i giovani specialisti

Napoli, 15 luglio 2010

Carlo Antonio Leone

AUTORI

Camaioni A.

U.O.C. ORL

Azienda Ospedaliera “San Giovanni – Addolorata” di Roma

Canzi P.

Dipartimento di Otorinolaringoiatria,

Università degli Studi e Fondazione IRCCS Policlinico San Matteo, Pavia

Cavicchi O.

Clinica Otorinolaringoiatrica

Università di Bologna – Policlinico S.Orsola

Chiesa F.

Divisione di Chirurgia Cervico-Facciale

Istituto Europeo di Oncologia, Milano

De Benedetto L.

Divisione di Chirurgia Cervico-Facciale

Istituto Europeo di Oncologia, Milano

Mele V.

S.O.C. ORL,

Az. Ospedaliero-Universitaria “S. Maria della Misericordia” Udine

Palma S.

S.O.C. ORL,

Az. Ospedaliero-Universitaria “S. Maria della Misericordia” Udine

Piemonte M.

S.O.C. ORL,

Az. Ospedaliero-Universitaria “S. Maria della Misericordia” Udine

Rinaldi Ceroni A.

Clinica Otorinolaringoiatrica

Università di Bologna - Policlinico S.Orsola

Rugiu M.G.

S.O.C. ORL,

Az. Ospedaliero-Universitaria “S. Maria della Misericordia” Udine

Simone M.

U.O.C. ORL

Azienda Ospedaliera “San Giovanni – Addolorata” di Roma

Vicini C.

U.O. ORL e Chirurgia Cervico-Facciale,
Ospedale G.B. Morgagni – L. Pierantoni, Forlì

Vidale C.

Documentalista Biomedico,
Az. Ospedaliero-Universitaria “S. Maria della Misericordia” Udine

Viti C.

U.O.C. ORL
Azienda Ospedaliera “San Giovanni – Addolorata” di Roma

INTRODUZIONE

M. Piemonte

La conoscenza scientifica e l'approfondimento culturale hanno rappresentato da sempre elementi irrinunciabili della professione medica: per secoli tali elementi sono stati devoluti esclusivamente alla trasmissione orale, alla tradizione e diffusione scritta, dapprima con libri e testi, poi anche con le riviste scientifiche, e ovviamente all'empirismo pratico dei professionisti.

Questo tipo di tradizione culturale presentava certamente rilevante efficacia, in considerazione dei tempi e in concomitanza con l'assenza di valide alternative, ma trovava altresì limitazioni e difetti che oggi non sarebbero più considerati accettabili. Tra questi possiamo ricordare:

- *la difficoltà di accedere alle fonti scritte ed orali, che non sempre erano agevolmente raggiungibili per tutti;*
- *i tempi necessari alla stesura scritta, alla stampa e alla diffusione delle informazioni scientifiche, che in alcuni casi le rendevano obsolete già alla loro uscita o superate anzitempo;*
- *la difficoltà di eseguire ricerche di dati e bibliografie complete, in considerazione della frequente frammentarietà e dispersione delle fonti di informazione e la difficoltà di poterle mettere a confronto razionalmente e in modo ottimale;*
- *la mancanza di modalità veramente "scientifiche" e riconosciute per verificare l'attendibilità e la significatività delle informazioni in esse contenute;*
- *la frequente "selettività" dell'accesso alle fonti di informazione medica, raggiungibili spesso sulla base di casuale o fortuita vicinanza a biblioteche, ospedali, università o altre strutture di riferimento o consultabili solo con costi economici non sempre facilmente sostenibili;*
- *l'impossibilità di poter contare su fonti di riproduzione o di immagazzinamento delle informazioni diverse dalla trascrizione manuale o dall'acquisizione mnemonica.*

A ciò si aggiunga che per molti secoli il sapere scientifico, e in particolare quello medico, è stato riservato ai soli professionisti addetti ai lavori (medici, farmacisti, ecc.), con preclusione pressoché completa per l'uomo "comune".

Molti di noi ricordano con particolare riconoscenza l'introduzione nell'uso pratico di alcune preziose innovazioni tecniche e culturali che fino agli

anni '80 hanno accompagnato l'impegno professionale di studio e di ricerca del medico, quali in particolare:

- la possibilità di copiare, conservare, appuntare e rileggere "ad libitum" i testi scientifici (articoli in primo luogo) non di propria proprietà grazie alla macchine fotocopiatrici;
- la possibilità di ricevere e trasmettere via fax informazioni e testi in tempo reale;
- l'utilizzo oculato, non sempre agevole ed immediato, ma certamente molto proficuo delle prime "banche dati" cartacee rappresentate dagli indici bibliografici medici a stampa quali Index Medicus, Excerpta Medica, Current Contents, ecc.

Su questi mezzi, ancor oggi in parte utilizzati seppure con minore intensità, l'accesso e la circolazione delle informazioni scientifiche e della cultura medica hanno potuto fare sicuro affidamento per alcuni lustri, con importanti vantaggi rispetto agli anni antecedenti.

La vera rivoluzione culturale in questo campo si è peraltro realizzata, in tempi brevissimi, negli ultimi trent'anni, con l'invenzione, l'introduzione nell'uso pratico e la diffusione ormai ubiquitaria delle tecnologie informatiche e della Rete (Internet), che hanno modificato radicalmente – nel bene e nel male – il settore dell'informazione e della cultura anche in campo medico.

Queste acquisizioni hanno di fatto determinato un salto che potremmo definire "quantico" nelle possibilità di diffusione, ricerca e acquisizione di informazioni in tutti campi, ma soprattutto in quei settori (come quello medico) in cui l'evoluzione e l'innovazione sono particolarmente veloci e la produzione culturale è particolarmente vasta.

La tecnologia informatica e la Rete hanno di colpo azzerato gran parte dei problemi preesistenti di ricerca e acquisizione delle informazioni scientifiche in campo medico, creando peraltro alcuni problemi di nuova e non indifferente incidenza.

In particolare il mondo dell'informazione e della cultura medica nell'età informatica è caratterizzato, al contrario dei tempi passati, da:

- universalità dell'informazione, in quanto ormai da qualunque punto del mondo è possibile accedere a tutto il "corpus" scientifico contenuto in Rete;
- facilità di accesso, in quanto non sono più necessarie spese rilevanti o la vicinanza di attrezzature/strutture dedicate per accedere a tali informazioni;
- immediatezza e tempestività della diffusione delle informazioni, che oggi vengono immesse in Rete in tempo reale non appena

si rendono disponibili e immediatamente sono raggiungibili da qualunque utente;

- completezza delle informazioni, in quanto ogni utente può accedere in qualunque luogo (salvo limitazioni commerciali o di legge) a tutte le informazioni contenute in Rete, alla pari di ogni altro utente del mondo;
- facilità di acquisizione, riproduzione e/o memorizzazione con vario mezzo (memoria informatica, trasferimento a stampa, ecc.) e talora rielaborazione ("on line", "off line") delle informazioni acquisite;
- possibilità di confronto rapido o addirittura immediato, con altri studiosi e professionisti, attraverso gli stessi canali informatici.

A questo impressionante progresso si è associata un'importante innovazione, rappresentata dall'introduzione nell'uso pratico di "motori di ricerca" generali o dedicati e di "banche dati" che non solo facilitano l'attività di ricerca di dati scientifici e bibliografie, ma soprattutto consentono una completezza, un approfondimento e una specificità di ricerca del tutto impossibili e impensabili in altri tempi.

L'attivazione di canali informatici di accesso alla cultura scientifica (e medica in particolare) ha peraltro conseguito un ulteriore, rilevante risultato, che potremmo riassumere nella "libertà" e nella "globalizzazione" dell'informazione: il "corpus" scientifico medico non è più riservato ai professionisti e a pochi altri fortunati, ma è diventato di colpo disponibile all'intera popolazione che sappia usare un computer.

È peraltro evidente che questa liberalizzazione e globalizzazione dell'accesso all'informazione sanitaria solleva importanti problemi sia per quanto riguarda il personale medico e sanitario in generale, sia per i cittadini privi di una formazione e di una cultura medica di base e specifica.

Tali problemi, con diversa incidenza, rilevanza e modalità di espressione per professionisti e per comuni cittadini, sono rappresentati principalmente da:

- numerosità delle fonti;
- serietà e attendibilità delle fonti;
- significatività dei dati riportati;
- utilizzo delle informazioni.

La numerosità delle fonti è diretta conseguenza delle caratteristiche sopra citate di universalità, facilità di accesso, completezza delle informazioni e dell'estensivo utilizzo di motori di ricerca e banche dati.

A titolo esemplificativo, la ricerca nella Rete di tre semplici parole chiave, in italiano (I) o in inglese (E), quali "tonsillectomia" (I) o "tonsillectomy" (E),

“laringectomia” (I) o “laryngectomy”, “otite” (I) o “otitis” (E) su tre diversi motori di ricerca, due generali (Google in italiano, Google in Inglese) e uno dedicato (PubMed in inglese), ha dato esito (16/7/2010) al riscontro del seguente numero di voci:

- Tonsillectomia/tonsillectomy
- Google (I) 4.270 Google(E) 675.000 PubMed (E) 8.056
- Laringectomia/laryngectomy
- Google (I) 85.800 Google(E) 184.000 PubMed (E) 8.528
- Otite/Otitis
- Google (I) 1.090.000 Google(E) 2.160.000 PubMed (E) 27.240

Appare evidente l'impossibilità di destreggiarsi utilmente in una selva così numerosa di riferimenti informativi e quindi la necessità (e la capacità) di saper specificare e restringere i campi di ricerca per ridurre i riferimenti a numeri meglio gestibili e precisi.

Il corretto ed oculato utilizzo dei motori di ricerca e delle banche dati rappresenta pertanto il primo, indispensabile elemento per un corretto accesso alle fonti informative oggi a disposizione in via informatica.

La serietà e l'attendibilità delle fonti rappresenta il secondo problema di attualità in questo campo, in quanto le caratteristiche di libertà e universalità dell'accesso alla rete informatica (che rappresentano una delle più importanti e rivoluzionarie caratteristiche di questa innovazione tecnologica) non presentano oggi limitazioni significative né sistemi di controllo di qualità “alla fonte”.

Di fatto, qualunque informazione – giusta o sbagliata, vera o non verificata, dimostrata o puramente opinabile /fantasiosa – può essere impunemente inserita in Rete e resa accessibile a tutti senza alcun controllo o “caveat”. Questo espone il navigatore in Rete e l'utente di banche dati al rischio di dare per vera/dimostrata/accertata un'informazione che tale non è: certamente questo rischio è in genere minore per il professionista esperto di un argomento, che cerca per via informatica aggiornamenti, conferme o precisazioni in merito a notizie già note di base, rispetto al soggetto privo delle cognizioni di base o comunque di elementi di confronto critico.

La conoscenza e la ricerca delle fonti più attendibili (o, in negativo, di quelle poco affidabili) rappresentano momenti irrinunciabili per una corretta attività di ricerca scientifica e bibliografica in ambito medico.

L'elemento problematico più importante è tuttavia rappresentato dalla significatività dei dati riportati, in quanto non necessariamente dati sia pure provenienti da fonte attendibile e seria, correttamente riportati ed interpretati, raggiungono le caratteristiche di utilità e utilizzabilità necessarie sulla base di una reale significatività scientifica. È infatti esperienza

frequente il riscontro di lavori scientifici di fonti autorevoli, correttamente sviluppati e descritti, su casistiche valide e sovrapponibili ma con risultati contrastanti e non immediatamente interpretabili.

È in questo settore che l'attenzione del professionista deve essere particolarmente vivace e deve essere affiancata ad una specifica competenza tecnica, basata su comportamenti e indicazioni universalmente condivisi che vengono comunemente identificati nella dizione "analisi critica della letteratura".

Un tentativo di soluzione a questo problema è stato identificato nella "Evidence Based Medicine" (EBM), cioè nella medicina fondata sulle prove di efficacia di una pratica diagnostica o terapeutica o di un intervento. L'EBM è oggi ampiamente diffusa ed accettata in tutto il mondo, ma non mancano giudizi fortemente critici che lamentano almeno tre limiti di questa pratica:

- la EBM viene talora presentata come unica ed assoluta verità scientifica in molti settori (pur a fronte di limiti e difetti non del tutto irrilevanti);
- la EBM, concettualmente, tende a sostituire ogni altro tipo di fonte di conoscenza;
- non tutte le conoscenze scientifiche sono riconducibili o comunque descritte in ambito EBM.

Da queste premesse, anche tra i professionisti medici si sono attivate discussioni approfondite che talvolta, esulando dai più ristretti confini in ambito propriamente epistemologico, hanno esteso i loro interessi nel filosofico, fino a fare riferimento alle categorie Kantiane di "analisi" e "sintesi".

A questi concetti "filosofici", peraltro corrispondono produzioni scientifiche corrispondenti, tra le quali ad esempio possiamo ricordare le "systematic reviews" rispetto alle "narrative reviews".

Pertanto, appare oggi evidente che la EBM rappresenti certamente un punto di riferimento irrinunciabile per il medico in tutte le sue attività, ma non è sempre e comunque il solo accettabile.

Anche l'indirizzo applicativo, a fini di valutazione della significatività prognostica, diagnostica, terapeutica, eziologica e di altri aspetti clinici, può influire sull'interpretazione e sull'analisi critica delle informazioni, richiedendo modalità di esame differenti.

L'utilizzo delle informazioni è oggi il quarto e non ultimo problema in questo campo.

Da un lato i professionisti stessi devono imparare ad organizzare, selezionare ed utilizzare in modo proficuo e costruttivo, nell'ambito della propria

attività di assistenza, di didattica e di ricerca la gran mole di informazioni che pervengono ogni giorno da molteplici canali.

D'altro canto il professionista deve ormai confrontarsi ogni giorno, nella sua pratica quotidiana, con informazioni mediche (o, talvolta, simil-mediche o para-mediche) che il cittadino raccoglie da canali informativi talvolta sovrapponibili, ma che vengono filtrate ed interiorizzate dagli interessati in modo ben diverso, sulla base di conoscenze di base spesso incomplete, parziali o addirittura inesistenti, di interessi o interpretazioni personali impropri, di pregiudizi o opinioni non condivisibili, di aspettative e speranze umanamente comprensibili ma non per questo scientificamente accettabili.

Non a caso da diversi anni è invalso negli USA il modo di dire che per i cittadini il medico più consultato ed ascoltato è il "dr. Google"!

Né si può in questa sede tacere l'importanza che le fonti bibliografiche e le banche dati rivestono, per medici e cittadini, nell'ormai frequente contenzioso medico-legale e giudiziario, ove il riferimento a linee-guida e corrette prassi mediche (EBM in primo luogo) è ormai all'ordine del giorno.

In conclusione, la capacità di analisi critica di fonti, informazioni, dati ed esperienze, che da sempre ha caratterizzato nei secoli l'attività del medico nel contesto della sua meravigliosa e difficile professione, viene oggi richiesta con modalità e con caratteristiche nuove ma non certo difformi dalle sue tradizioni storiche e culturali.

La corretta valutazione critica delle evidenze scientifiche, ottenute mediante studi rigorosi e metodologie universalmente accettate, rappresenta oggi un imperativo nelle attività di studio, di assistenza, di didattica e di ricerca di ogni medico, indipendentemente dal suo ruolo e dalle sue competenze professionali.

Lo specialista Otorinolaringoiatra, da sempre attento alle innovazioni scientifiche e alle esigenze innovative del suo contesto professionale, non può quindi astrarsi neppure da questa evoluzione culturale e deve trovare nelle Società Scientifiche di riferimento (nello specifico SIO, AOOI e AUORL) ogni possibile supporto in chiave informativa e formativa.

Per questo motivo sembra quanto mai opportuno che un "manuale" pratico sulla metodologia clinica applicata trovi utile inserimento nella collana di Quaderni di Aggiornamento AOOI, il cui scopo deve essere non solo l'aggiornamento specialistico, ma anche l'aggiornamento professionale inteso nella sua globalità di competenze, responsabilità e "cultura".

Bibliografia scelta

1. Akobeng AK: Evidence based child health: Principles of evidence based medicine. Arch Dis Child 2005; 90: 837-840
2. Akobeng AK: Understanding systematic reviews and meta-analysis. Arch Dis Child 2005; 90: 845-848
3. Cook DJ, Mulrow CD, Haynes RB: Systematic reviews: synthesis of best evidence for clinical decisions. Ann Intern Med 1997; 126, 376-380
4. Dickersin K, Scherer R, Lefebvre C: Systematic reviews: identifying relevant studies for systematic reviews. BMJ 1994; 309: 1286-1291
5. Egger M, Juni P, Bartlett C, Hoenstein F, Sterne J: How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess 2003; 7:1
6. McAlister FA, Clark HD, van Walraven C: The medical review article revisited: has the science improved? Ann Intern Med 1999; 131: 947-951
7. McGovern DPB, Valori RM, Summerskill WSM (eds): Key topics in evidence based medicine. Oxford: BIOS Scient. Publ., 2001
8. Mickenautsch S: Systematic reviews, systematic error and the acquisition of clinical knowledge. BMC Med Res Methodol 2010; 10; 53
9. Montori WM, Wilczynski NL, Morgan D, Haynes RB: Systematic reviews: a cross-sectional study of location and citation counts. BMC Med 2003; 1:2
10. Ritchey T: Analysis and Synthesis. On Scientific Method – Based on a Study by Bernhard Riemann. Systems Res 1991; 8: 21-41
11. Stewart LA, Parmar MK: Meta-analysis of the literature or of individual patient data: is there a difference? Lancet 1993; 341: 418-422
12. Williams DDR, Garner J: The case against “the evidence”: a different perspective on evidence-based medicine. Br J Psychiatry 2002; 180: 8-12

INDICE

Presentazione	pag. 3
Autori	» 5
Introduzione M. Piemonte	» 7
Semplici concetti di statistica utili da sapere S. Palma	» 17
EBM: principi costitutivi, validità e limiti O. Cavicchi, A. Rinaldi Ceroni	» 27
Classificazione e tipologia dei lavori scientifici M.G. Rugiu	» 37
Analisi critica di un articolo scientifico Aspetti generali S. Palma, M. Piemonte	» 57
Analisi critica delle procedure diagnostiche. Indici di performance del test S. Palma	» 69
Analisi critica delle procedure terapeutiche S. Palma. V. Mele	» 93
Fattori di rischio, come interpretarli e come analizzarli C. Vicini, P. Canzi	» 111
Fattori prognostici, interpretazione dei dati oncologici e loro analisi F. Chiesa, L. De Benedetto	» 123
Validità, limiti e rischi della ricerca dati tramite banche elettroniche nella pratica clinica A. Camaioni , C. Viti , M. Simone	» 133

Ricerca ed analisi delle fonti tramite banche (dati) elettroniche

C. Vidale, S. Palma » 145

Appendice - La tabella 2 x 2 universale

S. Palma » 165

SEMPLICI CONCETTI DI STATISTICA UTILI DA SAPERE

S. Palma

Questo capitolo introduttivo non è indispensabile da leggere. Il lettore che non ha molta simpatia per concetti di matematica può saltare questa parte e passare alla lettura dei capitoli successivi senza che venga meno la finalità didattica del manuale.

Tuttavia riteniamo che questa parte sia propedeutica e utile sia nell'agevolazione della comprensione delle analisi più strettamente "quantitative" in termini numerici esposte nei capitoli successivi, sia per la comprensione di alcuni concetti statistici tanto nella lettura quanto nella stesura di un articolo scientifico.

Si ribadisce comunque che la finalità di questo manuale non è assolutamente riscrivere un testo di statistica, per la quale anzi si rimanda a testi ben più autorevoli.

Scale di misura

Per poter analizzare dei dati clinici e confrontarli è necessario adottare un sistema di misura, il sistema di misura ovviamente deve essere il più idoneo e fornire una dimensione reale dell'oggetto misurato. È noto a tutti che per misurare una lunghezza utilizzeremo il metro, se invece volessimo misurare una massa utilizzeremo una bilancia utilizzando come misura il chilo e i relativi multipli e sottomultipli. Se invece volessimo "misurare" i colori sicuramente non potremo utilizzare questi sistemi di misura, ma un altro sistema che possa essere universalmente riconosciuto tale anche da confrontare le diverse misure. Le scale di misurazione sono di 4 tipi: scala nominale, scale ordinale, scale ad intervalli, scale di rapporti.

a) *Scala nominale*. È il livello di misurazione più basso. Serve principalmente per identificare categorie dicotomiche o mutuamente esclusive, es. maschio- femmina, sano – malato, sopra o sotto i 65 anni, ecc. Ad ogni categoria si può assegnare convenzionalmente un numero es. 0 per una categoria e 1 per l'opposto, ma ovviamente non può esistere un termine intermedio che numericamente sarebbe 0,5, ma che nella realtà non esiste. Identificare due categorie con 0 e 1 è come se le chiamassimo con nomi propri (esempio Gianni e Giacomo).

b) Scala ordinale. Questo tipo di scala è utilizzata per frequenti osservazioni cliniche che possono essere ordinate su una scala secondo un criterio. Gli esempi più tipici sono i rilievi clinici sul paziente: es. peggiorato – invariato – migliorato – guarito, oppure sulla sintomatologia: assente–lieve – moderata – grave. Attenzione però: questo tipo di valutazione è *qualitativo* e non quantitativo. A volte per comodità descrittiva e di analisi la scala ordinale viene espressa con uno “score”: per tornare all’esempio della sintomatologia i quattro gradi possono essere espressi con i numeri 0 (= assente) – 1 (= lieve) – 2 (= moderato) – 3 (grave). Nell’ambito di ciascuna classe si suppone che tutti gli elementi siano uguali, ovvero che i paziente “gravi” abbiano sostanzialmente un quadro clinico sovrapponibile e che differiscano dai pazienti classificati con un score sintomatologico diverso, ovvero che stiano peggio di quelli con sintomatologia moderato e che siano ancora peggiori rispetto a quelli con sintomatologia lieve. Tuttavia in questo tipo di scala non è previsto che vi sia la stessa differenza tra le varie categorie, ovvero la differenza di sintomatologia tra i lievi e i moderati non è la stessa che esiste tra i moderati e i gravi. Di questo bisogna tenere presente quando si utilizza uno score sintomatologico numerico. Infatti se dal punto di vista matematico tra 1 e 2 esiste la stessa differenza che vi è tra 2 e 3, questo non vale quando si utilizza la scala ordinale, poiché in questo caso i numeri rappresentano solo una qualità e non una quantità.

c) Scala ad intervalli. Questo tipo di scala consente di ordinare le misure e a differenza della precedente la distanza tra due misure qualsiasi è la stessa, ovvero la differenza tra una misura di 1 e una di 2 è la stessa che esiste tra 2 e 3. Con questa scala si possono misurare *dati discreti* ovvero numeri reali, ma interi e non negativi, come ad esempio la registrazione di episodi faringo-tonsillitici acuti in un gruppo di pazienti. Un paziente può manifestare 2 episodi all’anno, un altro 3 e un altro ancora 4. Come si osserva la differenza tra i numeri rimane invariata, nessun paziente potrà avere 2,4 episodi faringo-tonsillitici e tanto meno -1 episodi all’anno. Questo tipo di scala consente di eseguire operazioni matematiche tra i dati; tuttavia i risultati dell’elaborazione non necessariamente saranno dati discreti; rimanendo nell’esempio di registrazione di episodi faringo-tonsillitici per anno potremmo avere 4 pazienti con i seguenti dati: 2, 2, 3, 3, la somma sarà 10 (dato discreto) e la media 2,5 (dato non discreto).

d) Scala di rapporti. Rappresenta il livello più alto di misurazione. Per questo tipo di scala esiste un vero punto zero, con questo tipo di scala ad esempio si misurano il peso e la statura, ovvero *dati continui*. Quando si utilizzano dati continui sono possibili valori frazionari. In ambito ORL dati continui si possono ottenere ad esempio misurando il volume delle cavità

nasali o le resistenze delle stesse al flusso aereo, oppure misurando la latenza delle onde nella registrazione dei potenziali evocati acustici.

Test parametrici e non parametrici. Concetto di media

I test *parametrici* fanno riferimento a test statistici che impiegano nella loro analisi la media e la deviazione standard. I dati che possono essere elaborati con questi test sono solo quelli misurati con le scale ad intervalli o di rapporti.

I test che non fanno riferimento a media e deviazione standard sono detti *non parametrici* e sono basati su frequenze, ranghi e percentili.

La media è un semplice calcolo matematico che consiste nella somma dei valori di ciascuna misura diviso per il numero di osservazioni o misure:

$$M = \frac{X_1 + X_2 + X_3 + \dots + X_n}{N \text{ misure}}$$

La media è quindi un semplice parametro per rappresentare una popolazione in funzione di un qualcosa misurabile con una scala ad intervalli o di rapporti. Tuttavia i singoli valori rilevati possono avere una dispersione molto variabile come illustrato nella fig. 1.

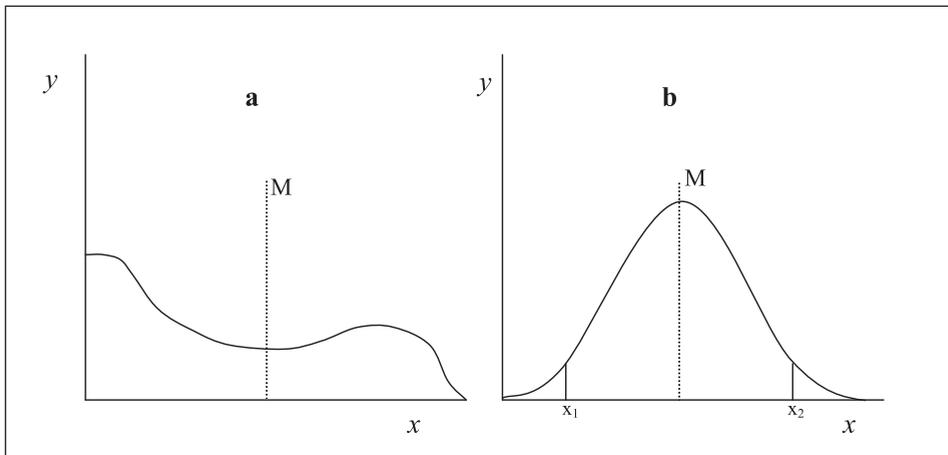


Fig. 1: rappresentazione schematica di possibili distribuzioni di misurazioni di valore x in funzione della frequenza dei singoli valori (asse y). Nella curva "a" la distribuzione è molto ampia e casuale, nella curva "b" la distribuzione è di tipo gaussiano (distribuzione "normale"). M indica il valore di x corrispondente alla media

Ipotizzando di misurare il peso di tutti i soggetti di una popolazione o di un campione di una popolazione si possono avere diverse distribuzioni dei valori. Ad esempio nella curva "a" della fig. 1 la distribuzione è molto ampia e casuale e il valore medio può coincidere con il valore rappresentato con meno frequenza (indicata sull'asse y). La curva "b" invece rappresenta una *distribuzione normale o gaussiana* in cui la maggior parte dei valori sono distribuiti intorno al valore medio. Questo tipo di curva è definito anche a "2 code". Per comprendere meglio questo aspetto è necessario introdurre il concetto di *deviazione standard*.

La misurazione di uno stesso parametro (esempio peso corporeo) in due popolazioni diverse può portare alla stessa media come schematicamente è riportato nella fig. 2. In questa figura le due curve, "a" e "b", hanno la stessa media, ma altezza diversa, ovvero la curva "a" è più stretta e alta, mentre la "b" è più bassa e larga indicando una maggior dispersione dei valori. Una possibile diversificazione delle due popolazioni può essere ottenuta considerando i singoli scarti, ovvero le differenze di ogni misurazione con la media. Tuttavia la semplice somma algebrica degli scarti (alcuni con valori negativi perché inferiori alla media, altri positivi perché superiori alla media) porterebbe inevitabilmente ad una somma nulla. Per ovviare a questo inconveniente si ricorre ad un semplice espediente matematico: si sommano i quadrati dei singoli scarti, per cui anche i valori negativi al quadrato diventano positivi. La somma dei quadrati degli scarti rapportata al

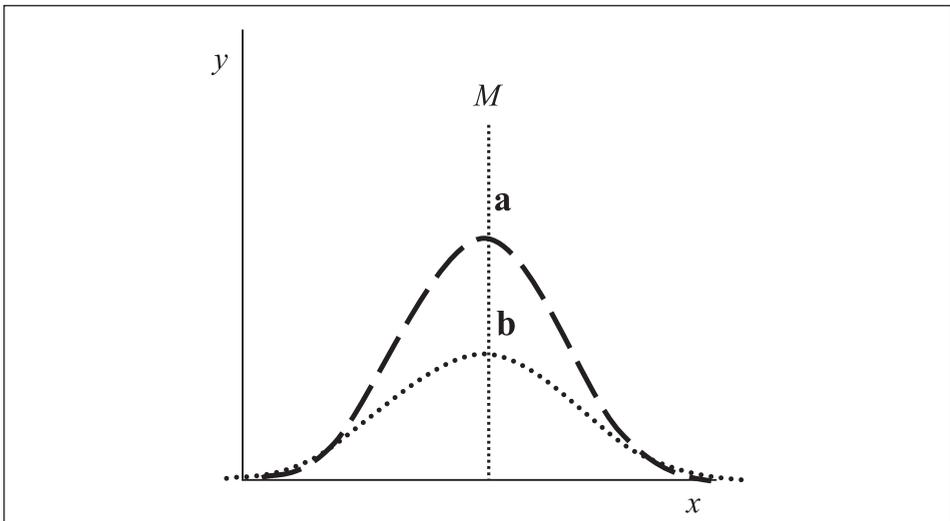


Fig. 2: due curve gaussiane con la stessa media, ma con deviazione standard diversa.

numero di osservazioni (o numero di elementi della popolazione) è detta *varianza* (σ^2).

$$\text{Varianza} = \frac{\sum (x_n - M)^2}{N}$$

La radice quadrata della varianza è la *deviazione standard* ($\sigma = \sqrt{\sigma^2}$). In una distribuzione normale il 95% delle misurazioni sono contenute in un range di valori pari alla media più o meno due volte la deviazione standard ($M \pm 2 \sigma$). Tornando alla fig. 1b ipotizziamo che i valori x_1 e x_2 corrispondano rispettivamente a $M - 2 \sigma$ e a $M + 2 \sigma$, tra questi 2 valori cadono il 95% delle osservazioni. Il 5% delle osservazioni è invece distribuito nelle “code” della curva ovvero “al di là” di $M \pm 2 \sigma$, in particolare il 2,5% è inferiore a x_1 , e il 2,5% è superiore a x_2 .

La media e la deviazione standard (d.s.) consentono di definire meglio la distribuzione delle osservazioni di una popolazione. Nel caso della fig. 2 le due popolazioni hanno sì la stessa media, ma differiscono per la deviazione standard, che sarà minore per la curva “a” e maggiore per la curva “b”. Quanto minore è la d.s. tanto più stretta sarà la curva gaussiana e minore sarà la dispersione dei dati intorno alla media. Una d.s. tendenzialmente elevata indicherà una maggiore dispersione dei dati.

Nei lavori scientifici il dato che più comunemente viene espresso con la media e d.s. è l’età dei pazienti arruolati nello studio. Si raccomanda di prestare attenzione al valore della d.s. perché dà una idea della distribuzione dei valori osservati, ricordando che il valore doppio della d.s. sommato e sottratto alla media dà gli estremi del range in cui sono distribuiti il 95% delle osservazioni.

Altri parametri che sono utili per descrivere i dati osservati in una popolazione sono la *moda* e la *mediana*. La moda è il valore osservato più volte, la mediana è esattamente il valore centrale delle misure raccolte. Esempio: poniamo di aver raccolto in un campione di 10 pazienti le seguenti misure secondo una scala ad intervalli: 3, 3, 4, 5, 5, 7, 8, 9, 9, 9. La media risulta 6,2, la d.s. 2,5 (il 95% dei valori è compreso tra $6,2 \pm (2,5 \times 2)$, ovvero tra 1,2 e 11,2, una dispersione piuttosto ampia non gaussiana), la moda è 9 (il valore più frequentemente espresso), la mediana è 6. Se le osservazioni sono indicate da un numero dispari la mediana corrisponderà al numero centrale, se invece sono indicate da un numero pari come è nell’esempio la mediana è la media dei due numeri centrali (nell’esempio 5 e 7, la cui media è $(5 + 7)/2 = 6$). In sintesi la mediana si calcola come $(N + 1)/2$, dove N è il numero di osservazioni disposte in ordine cre-

scenze. A volte media, mediana e moda quasi coincidono, come accade in una distribuzione normale, nell'esempio da noi riportato solo media e mediana quasi coincidono, mentre la moda è ben diversa. Quindi se si vuole dare una idea più precisa della distribuzione dei propri dati è più utile riportare oltre alla media e d.s. anche la mediana e la moda; se questi valori quasi coincidono allora è verosimile che la distribuzione sia di tipo normale e quindi su questi dati trova giustificazione l'applicazione dei test parametrici.

Concetto di percentile. La mediana indica un valore che consente di dividere in due parti le misure rilevate: 50% sono inferiori alla mediana e 50% invece superiori. Pertanto si può definire mediana anche come 50° percentile. Analogamente il 25° percentile indica il punto che delimita il quarto inferiore delle osservazioni e corrisponde all'osservazione $(N + 1)/4$. In linea generale il percentile di ordine p -esimo è l'osservazione che occupa la posizione numero $(N + 1)/(100/p)$.

Nella pratica i percentili più usati sono il 25° e il 75° che includono le osservazioni tra il quarto inferiore e il quarto superiore. Utilizzando quindi i percentili si possono avere ulteriori informazioni sulla distribuzione delle osservazioni. Nel semplice esempio riportato prima (10 osservazioni: 3, 3, 4, 5, 5, 7, 8, 9, 9, 9) il 25° percentile corrisponde a 3,75 (cade a $\frac{3}{4}$ tra 3 e 4), mentre il 75° a 9 (in quanto cade tra l'ottava e nona osservazione).

Concetto di rango. Il rango è il numero d'ordine di un valore in una serie ordinata dei valori stessi. Al valore più piccolo si attribuisce il rango più piccolo corrispondente a 1, al valore successivo più grande sarà attribuito il rango 2 e così via. In caso di 2 valori uguali il rango di entrambi sarà la media dei due ranghi consecutivi, per esempio nella sequenza dell'esempio precedente la 4^a e 5^a osservazione hanno entrambe il valore 5, entrambe avranno il rango 4,5 $[(4 + 5)/2]$. La distribuzione per ranghi si applica quando le osservazioni non hanno una distribuzione gaussiana o quando sono rilevate con scale ordinali. In questo caso l'analisi avverrà con test non parametrici.

Media ponderata. Talvolta in alcuni lavori scientifici si fa riferimento alla media ponderata. È un modo diverso di elaborare i dati osservati basandosi non solo sui singoli valori, ma anche sulla importanza (peso) che si attribuisce a questi valori. Dal punto di vista matematico la media ponderata P è espressa come rapporto tra somma dei singoli prodotti tra valore osservato e rispettivo peso e la somma dei singoli pesi:

$$P = \frac{x_1p_1 + x_2p_2 + \dots + x_np_n}{p_1 + p_2 + \dots + p_n}$$

frequenza (Hz)	250	500	1000	2000	3000	4000	6000	8000
dB	10	20	30	30	35	40	50	60
peso	1	2	3	2	1	1	1	1
dB x peso	10	40	90	60	35	40	50	60

Tab. I: esempio di soglia audiometrica a cui sono stati attribuiti pesi per importanza di frequenza.

Il valore del peso può essere assegnato arbitrariamente in base all'importanza che si ritiene possa avere una data rilevazione, importante è che nel denominatore siano sommati i pesi utilizzati e non compaia semplicemente il numero delle osservazioni. Ad esempio se volessimo eseguire una media ponderata di una soglia audiometrica, potremmo stabilire che le frequenze 500, 1000 e 2000 siano più rilevanti e attribuire a loro i seguenti pesi 2, 3 e 2.

Nella Tab. I a scopo esemplificativo abbiamo riportato dei valori di una possibile soglia determinata per via aerea di un orecchio. Se dovessimo calcolare la media aritmetica sommeremo i valori della prima riga, dividendoli per il numero delle osservazioni (frequenze) cioè 8; la media che otterremmo è di 34,375. Se invece volessimo eseguire una media ponderata con i pesi da noi assegnati, utilizzeremo la somma dei dati della terza riga dividendola per la somma dei pesi della seconda riga e otterremmo il valore di 32,1. Poiché nella comprensione delle parole le frequenze più acute incidono meno, la media ponderata dà un'idea migliore della performance uditiva del soggetto esaminato. In ambito medico quando le osservazioni non hanno un'uguale importanza sull'impatto clinico è bene differenziarle dando loro un peso diverso valorizzando quelle che possono essere più significative.

La media ponderata è utilizzata anche quando i dati sono suddivisi in diverse classi, esempio tipico è la suddivisione per fasce di età.

Se consideriamo come esempio i dati della Tab. II la media ponderata viene calcolata come segue. Per ogni fascia di età si considera il valore centrale, rispettivamente 15, 25, 35 e 45. Questi valori vengono moltiplicati per le rispettive frequenze: 15 x 6, 25 x 8, 35 x 7, 45 x 10; la somma

Fasce d'età	N.° soggetti
10 – 20	6
20 – 30	8
30 – 40	7
40 – 50	10

Tab. II: esempio di campione di soggetti suddiviso per fasce d'età

di questi prodotti (985) viene divisa per la somma delle frequenze (31). Il valore medio ponderato è di 31,77.

In sintesi. Quando si progetta uno studio è importante stabilire cosa misurare e come misurare. È importante sapere quale tipo di scala useremo, perché in base al tipo di scala usato otterremo dei dati che devono essere elaborati con i corretti test statistici. Quando possiamo ottenere dei dati con scale ad intervalli o di rapporti e la distribuzione dei dati è di tipo gaussiano, allora possiamo applicare test parametrici facendo riferimento a media e deviazione standard. Quando invece non vi è una distribuzione gaussiana o sono state impiegate scale ordinali o nominali, allora occorre impiegare test non parametrici.

Tabelle 2 x 2 o di contingenza

In ambito clinico sono molto utilizzate le tabelle 2 x 2, così definite in quanto costituite da due colonne e due righe (Tab. III). Questo tipo di tabelle, definite anche di contingenza, sono utilizzate per suddividere i dati ottenuti con scale nominali, quando la suddivisione è dicotomica, esempio tipico maschi e femmine. In questo caso la distribuzione dei dati avviene per frequenze e la scala è definita anche binominale. Questo tipo di tabella è quella che si impiega quando si utilizza il test statistico χ^2 , dove vengono incrociate due variabili una espressa nelle righe, l'altra nelle colonne. Un esempio semplice è suddividere soggetti biondi e bruni per sesso. Se nelle colonne si riporta il sesso (maschi in C1 e femmine in C2) e nelle righe i colori dei capelli (biondi in R1 e bruni in R2), le caselle centrali indicate con le lettere a, b, c e d, che definiscono propriamente la tabella 2 x 2, riporteranno le rispettive frequenze, ovvero nella casella "a" avremo il numero di soggetti maschi biondi, nella casella "b" il numero di femmine bionde, nella "c" il numero di maschi bruni e in "d" il numero delle femmine brune. La casella in basso a destra dove è indicato "n" riporta il totale del campione esaminato. Le somme delle colonne e delle righe riportano rispettivamente i totali di soggetti maschi e femmine e i totali dei soggetti biondi e bruni.

Nei capitoli successivi per la valutazione dell'impatto clinico dei risultati ricavati dai lavori scientifici (in particolare per gli effetti della terapia e per l'analisi dei test diagnostici) si utilizzeranno spesso questo tipo di tabelle. La Tab. IV ne è un esempio tipico. Il campione di pazienti in esame viene suddiviso secondo una variabile dicotomica: un gruppo possiede una certa caratteristica, esempio una determinata malattia, l'altro gruppo invece non possiede questa caratteristica, esempio

	C1	C2	Tot.
R1	a	b	a+b
R2	c	d	c+d
Tot.	a+c	b+d	n

Tab. III: esempio di tabella 2 x 2

esenti dalla malattia in esame. La seconda variabile espressa nelle righe è l'azione o l'osservazione dello sperimentatore. Nella prima riga sono indicati i soggetti che sono stati sottoposti ad un certo trattamento terapeutico di cui si vuole valutare l'efficacia, oppure sottoposti ad un certo test diagnostico risultato positivo, o ancora possono essere indicati i soggetti esposti ad un determinato fattore di rischio. Nella seconda riga invece sono indicati i soggetti non sottoposti al trattamento in esame (esempio gruppo placebo o controllo), oppure i soggetti risultati negativi al test in esame, oppure i soggetti non esposti al fattore di rischio che si sta studiando. Risulta facilmente evidente che il numero di soggetti indicati nella casella "a" sono quelli che hanno la malattia considerata e che sono stati sottoposti o esposti rispettivamente o ad una terapia o ad un fattore di rischio. Nella casella "d" invece sono indicati coloro che non hanno la malattia e che non hanno subito alcuna esposizione ad un trattamento terapeutico o fattore di rischio. Nel caso si voglia verificare l'efficacia di una terapia ci si aspetta che i soggetti nella casella "a" siano percentualmente pochi, mentre se il fattore di rischio (che può essere anche un trattamento farmacologico) è strettamente correlato alla malattia allora il numero percentuale in quella casella sarà tendenzialmente alto.

Per le varie esemplificazioni si rimanda ai capitoli successivi.

	M +	M -	Tot.
Esposti	a	b	a + b
Non esposti	c	d	c + d
Tot.	a + c	b + d	n

Tab. IV: tabella 2 x 2 per analisi clinica

Bibliografia

- Glantz SA. *Statistica per Discipline Biomediche*. V^a Edizione. Ed. McGraw-Hill, Milano 2003.
- Pagano M, Gauvreau K. *Biostatistica*. IIa Edizione. Gruppo Editoriale Idelson-Gnocchi, Napoli 2003.
- Daniel WW. *Biostatistica: concetti di base per l'analisi statistica delle scienze dell'area medico-sanitaria*. EdiSES S.r.l., Napoli 1996.

EBM: PRINCIPI COSTITUTIVI, VALIDITÀ E LIMITI

O. Cavicchi, A. Rinaldi Ceroni

Introduzione

La Evidence-based Medicine (EBM) è un movimento culturale che si è progressivamente diffuso a livello internazionale, favorito da alcuni fenomeni che hanno contribuito ad una crisi dei modelli tradizionali della medicina. La nascita di tale movimento culturale appare come uno spartiacque definitivo tra la medicina empirica spesso applicata senza prove e la medicina basata sulle prove di efficacia di un intervento o di una procedura diagnostica. La definizione attualmente più accreditata di EBM è fornita da David Sackett¹: **“la EBM costituisce un approccio alla pratica clinica dove le decisioni sono la conseguenza dell’integrazione tra l’esperienza del medico e l’utilizzo coscienzioso e giudizioso delle migliori evidenze scientifiche disponibili, mediate dalle preferenze del paziente”**. La EBM, prevede pertanto un processo di autoapprendimento in cui l’assistenza al paziente individuale stimola alla ricerca in letteratura di informazioni valide, rilevanti ed aggiornate che consentono al medico di “colmare” i gap di conoscenza.

Principi costitutivi

I primi segnali di cambiamento del modo di interpretare la professione medica risalgono al 1830 quando Pierre Charles Alexandre Louis, a Parigi, promosse la “Médecine d’Observation”, un movimento secondo cui i medici, piuttosto che affidarsi esclusivamente all’empirismo e alle speculazioni sulle cause di malattia, dovevano agire sulla base di ampie serie sperimentali, capaci di fornire una stima quantitativa degli effetti terapeutici. L’autore poneva così le fondamenta della EBM anche se bisognerà aspettare molti anni per l’avvento del vero fondatore del pensiero di questo movimento culturale.

Nel **1972** Archibald Cochrane², un epidemiologo inglese, in un libro che ha lasciato una traccia profonda nella storia della medicina, sostenendo che i risultati della ricerca avevano un impatto molto limitato sulla pratica clinica scriveva: **“è causa di grande preoccupazione constatare come la professione medica non abbia saputo organizzare un sistema**

in grado di rendere disponibili, e costantemente aggiornate, delle revisioni critiche sugli effetti dell'assistenza sanitaria". In quegli anni A. Cochrane, inoltre, già consapevole della limitatezza delle risorse economiche e della crisi dei Sistemi Sanitari, suggeriva di rendere disponibili a tutti i pazienti solo gli interventi sanitari di documentata efficacia. Emergevano in questo modo già due problemi fondamentali che daranno poi il via all'affermazione del movimento culturale della medicina basata sulle evidenze. Il primo è legato all'enorme mole di materiale bibliografico che rendeva arduo se non impossibile un aggiornamento proficuo del medico. Il secondo è l'impossibilità economica da parte dei Sistemi Sanitari di offrire ai pazienti tutte le prestazioni a prescindere dalla loro efficacia. Nel suo lavoro l'Autore enfatizzava due concetti che diventeranno il core della EBM: il concetto di efficacia intesa come il fare le cose giuste e possibilmente bene e il concetto di efficienza intesa come valutazione del rapporto tra risorse impiegate e risultati ottenuti (in termini di salute) e tra costi e benefici.

Un altro anno fondamentale nella storia della EBM fu il 1981 in cui, alcuni ricercatori Canadesi appartenenti alla McMaster School,³ partendo dalle criticità espresse da Cochrane in merito all'aggiornamento bibliografico del medico pubblicarono una serie consecutiva di lavori scientifici fondamentali nella storia della EBM intitolati "How to read Clinical Journals", nel quale descrivevano le strategie di approccio critico alla letteratura biomedica.

Altra data importante è il 1993, anno in cui viene fondata la Cochrane Collaboration⁴, network internazionale costruito per preparare, aggiornare e disseminare in assoluta autonomia revisioni sistematiche degli studi clinici controllati con criteri di rigosità fondati sulla statistica applicata alla medicina e sull'eliminazione di qualsiasi conflitto di interesse. Il lavoro dei gruppi di revisione della Cochrane Collaboration viene distribuito mediante la Cochrane library, una pubblicazione elettronica aggiornata ogni tre mesi.

Nel 1994 Shaughnessy e Coll.⁵ coniarono l'acronimo di POEMs (Patient Oriented Evidence that Matters) come risposta alla critica che già emergeva all'EBM cioè quella di essere molto distante dalla pratica clinica quotidiana. Secondo l'Autore le caratteristiche che dovevano avere le evidenze per essere clinicamente rilevanti erano le seguenti: essere utili per la attività clinica e riguardare interventi sanitari su end-point clinicamente significativi per il paziente (mortalità, morbidità, qualità della vita). Ben presto l'EBM ha destato anche l'interesse dei Sistemi Sanitari i quali vedevano in questo movimento culturale la possibilità di estendere le metodologie di questa dottrina alla pianificazione della salute delle popolazioni e al contenimento dei costi sempre più importanti. Con la pubbli-

cazione del libro di Muir Gray⁶ compare così il termine Evidence-based Health Care (EBHC). Secondo l'autore, l'EBHC prevede la descrizione esplicita delle modalità per programmare l'assistenza sanitaria sostenendo che è necessario ricercare sistematicamente, valutare e rendere disponibili le migliori evidenze scientifiche, quali prove di efficacia degli interventi sanitari per pianificare le decisioni (e gli investimenti) che riguardano la salute di una popolazione⁷.

Nel 2005 si diffonde poi il termine di Evidence-based Practice a sottolineare in maniera univoca la "pratica professionale basata sulle evidenze scientifiche". Il termine viene coniato da un gruppo di esperti e pubblicato nel Sicily Statement on Evidence Based Practice⁸. Le conclusioni di tale lavoro sono che tutti i professionisti della sanità devono:

- possedere uno spirito critico sia nei confronti della propria pratica professionale, sia delle evidenze scientifiche;
- essere capaci di ricercare, valutare e applicare le migliori evidenze scientifiche (EBP core-curriculum);
- essere disponibili ad implementare linee guida e percorsi assistenziali.

In assenza di queste conoscenze/competenze/attitudini, è impossibile per le organizzazioni sanitarie fornire "la migliore assistenza, basata sulle migliori evidenze"⁶.

Validità

Ci sono stati indubbiamente dei fattori che hanno contribuito alla crisi dei modelli tradizionali di aggiornamento ed espletamento della pratica medica.

Prima della promozione ed affermazione dell'EBM l'aggiornamento del medico era affidato a corsi e congressi che, nella maggioranza dei casi, fornivano informazioni, ma non conoscenze. Spesso l'aggiornamento e la soluzione dei propri dubbi o quesiti veniva riposta nelle mani dei colleghi più esperti. Inoltre la maggior parte delle informazioni derivava poi dalle ditte farmaceutiche che svolgevano e svolgono tutt'ora la maggior parte della ricerca esponendo così il medico ad un importante conflitto di interessi. L'aggiornamento mediante i libri di testo diventava presto obsoleto considerando il velocissimo progredire delle ricerche e delle conseguenti pubblicazioni.⁹

La crescita esponenziale dell'informazione biomedica, sia per volume che per complessità, rendeva perciò sempre più arduo l'aggiornamento professionale costante. Per il singolo medico diventava impossibile,

anche per questioni di tempo, districarsi all'interno di una mole enorme di pubblicazioni scientifiche senza sapere, nella maggioranza dei casi, quali lavori scientifici potessero essere veramente utili per il suo aggiornamento o per il miglior approccio ad ogni singolo paziente. L'approccio critico alla letteratura biomedica rappresenta sicuramente la componente più intima dell'EBM. Infatti la sorveglianza periodica della letteratura, integrata all'approccio critico, consente di identificare quali risultati della ricerca possono essere incorporati nella pratica clinica⁷ differenziando i lavori scientifici in base alla tipologia e alla loro qualità. Nasce così la cosiddetta "Piramide delle Evidenze" secondo la quale i lavori scientifici vengono classificati secondo uno schema che prevede alla base gli studi preliminari di base, quindi quelli che esprimono opinioni di esperti, i Case Report, la Serie di Casi, gli studi Caso-Controllo, gli studi di Coorte ed all'apice i Trial Clinici Randomizzati (RCTs) e le Metanalisi di RCTs¹⁰. I RCTs e le metanalisi di RCTs sono ormai riconosciuti universalmente come il metodo di analisi dotato di maggior valore scientifico rispetto alle review tradizionali in quanto caratterizzati da minori possibilità di errore metodologico (bias).¹²

L'affermarsi dell'EBM è stato indubbiamente agevolato dall'ausilio delle tecnologie telematiche che facilitano l'aggiornamento periodico delle conoscenze mediche consentendo di avere in tempo reale un contatto con siti web nazionali ed internazionali contenenti banche dati, Trial RCTs, Linee Guida (LG), ma soprattutto metanalisi di RCTs che sono un riassunto pratico e veloce di tutta la bibliografia indicizzata e accreditata su un argomento.

Altro miglioramento apportato dalla introduzione della Medicina basata sulle Evidenze è senz'altro la riduzione del divario tra i risultati della ricerca e la loro applicazione sul paziente nella pratica quotidiana. Questa divaricazione riconosceva fattori interni ed esterni alla professione medica. Tra quelli interni si può ricordare l'ampia variabilità della pratica professionale, la riluttanza da parte del medico a cambiare le proprie convinzioni con il conseguente e persistente utilizzo empirico di trattamenti inefficaci, la mancanza di aggiornamento o la difficoltà di aggiornarsi. Tra i fattori esterni va ricordata la presenza di influenze nell'aggiornamento del medico da parte delle ditte farmaceutiche e tecnologiche e anche la moda, intesa come il seguire pratiche cliniche ormai entrate nella routine, ma non sempre completamente provate. Attualmente la metodologia EBM dovrebbe consentire di valutare velocemente l'efficacia di una nuova procedura sia terapeutica che diagnostica e, volendo, anche di misurarne l'impatto e la convenienza economica.

Il successo dell'EBM è stato poi promosso dall'interesse che questa ha provocato nei Sistemi Sanitari che hanno intravisto in questo movimento

culturale un tentativo di soluzione della loro crisi economica determinata dalla crescita della domanda e dall'aumento dei costi dell'assistenza. L'EBM è diventata perciò anche un metodo da utilizzare per razionalizzare la spesa sanitaria promuovendo solo ciò che si è confermato essere efficace. La diretta conseguenza di tale fenomeno è stato il proliferare di LG e l'istituzione del Governo Clinico all'interno delle Aziende Sanitarie che hanno come fine, sicuramente non marginale, il mantenimento dei costi.

Le LG secondo la definizione più accreditata sono **“raccomandazioni sviluppate in modo sistematico per assistere medici e pazienti nelle decisioni sulla gestione appropriata di specifiche condizioni cliniche” (PNLG-glossario)**. La produzione delle LG è regolata da una rigida normativa, non sempre seguita, emessa dalla Convention di Atlanta del 1992¹³ e di seguito riportata:

- validità: cioè una volta applicata deve portare al beneficio atteso (salute dei pazienti/economico);
- riproducibilità: usando lo stesso metodo conduce esperti diversi ad ottenere gli stessi risultati;
- rappresentatività: interessare tutti i livelli professionali coinvolti;
- applicabilità;
- flessibilità: cioè chiarire tutte le circostanze che fanno eccezione;
- chiarezza;
- contenere tutto il percorso metodologico seguito per la produzione e validazione;
- deve essere sempre chiara la qualità delle evidenze su cui si fondano le raccomandazioni come conseguenza di una attenta e critica analisi della letteratura biomedica;
- periodicamente aggiornata.

Purtroppo le LG hanno dimostrato di avere anche degli importanti limiti dovuti principalmente alla notevole variabilità metodologica utilizzata per la loro elaborazione. Un esempio emblematico di questo fenomeno è senz'altro il proliferare delle griglie di lettura della evidenza scientifica che rispecchiano il tentativo di adeguarle alle proprie necessità. Infatti non vi è ancora un modo unitario di considerare la metodologia di valutazione complessiva della evidenza scientifica soprattutto sul livello della prova e sul grado di raccomandazione a tal punto che in letteratura esistono almeno 4 o 5 griglie di lettura

La griglia di lettura più semplice ed applicabile è quella condivisa da Agenzie come AHQR (US Agency for Healthcare Quality and Research), SIGN (Scottish Intercollegiate Guidelines Network), NICE (National Institute for Clinical Excellence (Tab.1).

Evidenza	Livello della prova	Grado della Raccomandazione	
Metanalisi di RCTs	Ia	A	Forte
Almeno 1 RCTs	Ib		
Almeno 1 studio clinico ben condotto senza randomizzazione	IIa	B	Discreta
Almeno 1 altro tipo di studio clinico ben disegnato quasi sperimentale	IIb		
Almeno 1 studio clinico ben disegnato non sperimentale	III		
Opinioni di comitati di esperti o esperienze di autorità riconosciute	IV	C	Debole

Tab. 1

Un altro limite delle LG è probabilmente quello di avere inasprito la crisi del rapporto tra autonomia professionale e controllo esterno; infatti pur essendo documenti elastici rappresentano pur sempre un riferimento tra medico e paziente dal quale può essere difficile e pericoloso discostarsi soprattutto quando la raccomandazione ha un grado elevato.¹⁴

Altro fattore limitante le LG è la loro scarsa implementazione ed aggiornamento e talora l'inapplicabilità che le rende di scarso impatto clinico nella pratica quotidiana.

Limiti

Il limite principale della EBM è sicuramente determinato dal fatto che essa non può essere applicata a tutti i campi della medicina. Esistono infatti dei settori in cui questa fatica ad affermarsi per una molteplice serie di motivi.

Per esempio i criteri dell'EBM non sembrano applicabili tali e quali alla chirurgia (EBS: Evidence Based Surgery), dove il cosiddetto "fattore chirurgico", fortemente condizionato dalla manualità e dall'esperienza del professionista e dalla strumentazione sempre più innovativa, risulta determinante man mano che le procedure diventano più complesse e la randomizzazione, che costituisce il nucleo dell'evidenza scientifica di qualità diventa sempre più difficile da realizzarsi per motivi anche e

soprattutto di natura metodologica e morale. La possibilità di applicare l'EBM diventa perciò minore quanto più è complessa la procedura. La difficoltà di applicazione dell'EBM alle procedure complesse si associa poi alla presenza di aree (aree grigie) della conoscenza medica dove non esiste ricerca di buona qualità, ma solo incertezza sull'efficacia di un intervento diagnostico-terapeutico e/o delle sue alternative.

Poichè l'interesse dei professionisti verso l'EBM viene condizionato dalla disomogenea distribuzione delle evidenze di buona qualità, la riduzione di tali aree grigie deve essere un punto di riferimento nella pianificazione della ricerca, al fine di ampliare progressivamente le evidenze necessarie alla pratica clinica⁶.

Anche gli RCTs che sono considerati il metodo di studio più affidabile per ottenere evidenze sulla efficacia degli interventi sanitari, non sono completamente immuni da critiche secondo gli esperti che vedono in questi una difficoltà di adattare i loro risultati al paziente preso individualmente. Un altro problema sorge frequentemente a livello di pianificazione dei RCTs soprattutto nella elaborazione dei criteri di inclusione e di esclusione. Gli RCTs vengono generalmente condotti su popolazioni selezionate ed omogenee, escludendo i pazienti "complessi" (comorbidità, anziani), le donne, i bambini, che rischiano di compromettere la validità interna dello studio. Per esempio se il ricercatore accetta di includere solamente persone con una certa cultura, capaci di leggere e di capire il linguaggio del consenso informato, che parlano correttamente la lingua italiana seleziona pazienti di un certo livello sociale. Ebbene i risultati dello studio avranno la massima affidabilità solo per quella popolazione studiata e con difficoltà potranno essere trasferiti sulla intera popolazione⁶. Esistono poi alcuni problemi legati sia agli end-point selezionati dagli sperimentatori che spesso non sono clinicamente significativi sia ai metodi utilizzati per riportare i risultati utilizzando accorgimenti statistici per enfatizzare l'efficacia dei trattamenti.

Tra i limiti estrinseci della applicazione della EBM va riportata anche la frequente scarsa attitudine del medico, soprattutto se considerato già esperto, a mettere in discussione le proprie conoscenze generando una vera e propria resistenza al cambiamento.

L'EBM e l'otorinolaringoiatria

In ambito Otorinolaringoiatrico l'applicazione delle metodologia EBM non ha la stessa diffusione di altri campi della Medicina. Ciò è dovuto principalmente al fatto che l'Otorinolaringoiatria è un branca "plurispecialistica" prevalentemente chirurgica per la quale, come precedentemente accen-

nato a proposito di EBS, risulta particolarmente indaginosa, ed in certi casi praticamente impossibile, l'applicazione dei criteri di inclusione ed esclusione nei RCTs. La difficoltà di giungere a conclusioni definitive e la diffusione sempre maggiore di problematiche medico-legali dà luogo a discussioni come quelle recentemente sollevate da Motta e Coll. alla pubblicazione della Linea Guida Nazionale (2008) a proposito di appropriatezza e sicurezza degli interventi di tonsillectomia e adenoidectomia. È inoltre indubbio che le nostre conoscenze di base attuali sono, in molti casi, limitate, concedendo così ampio spazio a provvedimenti clinici di tipo empirico e tradizionale che non si adattano a questa metodologia. Ciò avviene, per esempio, nelle ipoacusie improvvise dove nessuna delle terapie proposte in letteratura ha mai dimostrato una efficacia secondo i parametri EBM. D'altra parte bisogna sottolineare che esiste anche un ritardo della specialità dovuto ad una ancora scarsa sensibilizzazione su questo tema. A tal proposito l'articolo di Sharp e Coll.¹⁰, sullo stato delle conoscenze in Otorinolaringoiatria, è sufficientemente esplicativo. Infatti l'Autore riporta che nell'anno 2007 nel Medline erano presenti 250.000 pubblicazioni ORL; tra queste le revisioni sistematiche e i RCTS erano rispettivamente 826 (0,37%) e 4571(2,7%). Al contrario i case report e i lavori non classificabili secondo la metodologia EBM (reviews narrative, editoriali, trial non randomizzati) erano di gran lunga numericamente superiori rispettivamente 44.290 (20,1%) e 140523 (63%).

Bibliografia

1. Sackett DL, Rosenberg WMC, Muir Gray JA, Brian Hayness R, Scott Richardson W: Evidence based medicine: what it is and what isn't. *BMJ* 1996; 312:71-2
2. Cochrane AL: Effectiveness & efficiency: Random reflections on Health Services, The Royal Society of Medicine Press, 2004
3. Haynes RB, McKibbon KA, Fitzgerald D, Guyatt GH, Walker CJ, Sackett DL: How to read Clinical Journals. *Ann Intern Med* 1986; 105:309-12
4. <http://www.cochrane.org/>
5. Shaughnessy AF, Slawson DC, Bennet JH. Becoming an information master: a guidebook in the medical information jungle. *J Fam Pract* 1994; 39: 489-99
6. Muir Gray JA: Churchill Livingstone, 1997
7. <http://www.gimbe.org/eb/valutazione.asp>
8. Dawes M, Summerskill W, Glasziou P, Cartabellotta A, Martin J, Hopayan K, Porzolt F, Buris A, Osborne: Sicily Statement on Evidence – Based Practice *BMC Medical Education* 2005; 5:1186- 1472
9. Bryan-Brown CW, Drakup K *Am. J Crit Care*, 2004; 13:10-12
10. Sharp S, Richard J, Harvey MB, Martin J, Burton MA: The Epistemology of otolaryngology- head and neck surgery: A scientific evaluation of the knowledge base. *Otolaryngol Head Neck Surg* 2007; 137:61-65
11. CEBM- Oxford Centre for Evidence Based Medicine - Levels of Evidence and Grades of Recommendation
12. Burton MJ: Evidence-based medicine and otolaryngology-HNS: passing fashion or permanent solution. *Otolaryngol Head and Neck Surg* 2007; 137:47-51
13. Guidelines for Clinical Practice: from Development to Use, 1992 Insitute of Medicine
14. Grilli R: Linee Guida CARE 2003; 4:27-30

CLASSIFICAZIONE E TIPOLOGIA DEI LAVORI SCIENTIFICI

M.G. Rugiu

Introduzione

Obiettivo di questo capitolo è fornire una definizione generale delle principali tipologie di articoli scientifici pubblicati da riviste medico-scientifiche, basandosi sui criteri classificativi più comunemente utilizzati. Verranno descritti gli aspetti principali che caratterizzano i diversi tipi di pubblicazioni scientifiche, soprattutto per quanto riguarda la categoria dei *major papers*. Successivamente verranno esposti alcuni suggerimenti di carattere pratico su come scrivere correttamente un lavoro scientifico. Infine, chiuderanno il capitolo alcune annotazioni riguardanti le metodiche più comunemente utilizzate per la valutazione della qualità delle pubblicazioni scientifiche (Peer Review, Impact Factor), sottolineandone da una parte il loro indiscusso valore e dall'altra i loro principali limiti di impiego.

Esistono diversi criteri di classificazione dei lavori scientifici di argomento biomedico. Il più semplice prende in considerazione l'argomento di ricerca trattato nell'articolo; in base a questo criterio distinguiamo principalmente sei gruppi di pubblicazioni: terapia, diagnosi, eziologia, prognosi, qualità, economia. I primi quattro gruppi sono senza dubbio i più interessanti dal punto di vista clinico, e vedremo in seguito quali siano per ognuna di queste categorie i tipi di studio più appropriati.

Un altro tipo di classificazione degli articoli scientifici prevede una suddivisione in tre fondamentali categorie: i *major papers*, i *minor papers* e gli *invited papers*. Le ultime due categorie verranno brevemente trattate in seguito, mentre per il momento soffermiamo la nostra attenzione sui *major papers*, che rappresentano la categoria di pubblicazioni scientifiche di gran lunga più complesse e significative. Molto schematicamente i *major papers* possono essere distinti in due tipologie: gli studi primari (*original article* o *original research*) e gli studi secondari. Gli studi primari riportano i risultati di ricerche eseguite personalmente dagli autori, e possono essere a loro volta di tipo sperimentale o non sperimentale o osservazionale, questi ultimi sono caratterizzati dal fatto che il ricercatore indaga e riferisce su cosa sta accadendo, senza poter manipolare la distribuzione del fattore in studio, al contrario di quanto si verifica negli studi sperimentali. Gli studi osservazionali presentano due sottocatego-

rie: gli studi descrittivi che comprendono lo studio eziologico e lo studio trasversale (*cross-sectional study*) e gli studi analitici che comprendono lo studio di coorte (*cohort study*) e lo studio caso-controllo (*case-control study*).

La caratteristica fondamentale degli studi secondari è di essere basati sulla rianalisi dei dati ricavati da ricerche scientifiche eseguite da altri; essi sono rappresentati essenzialmente dalle revisioni sistematiche (*systematic review*) e dalle metanalisi, oltre ad articoli particolari, quali ad esempio quelli che utilizzano i risultati di studi primari per formulare analisi decisionali o analisi economiche (analisi costo/beneficio, costo/efficacia).

È interessante sottolineare fin da ora che le revisioni sistematiche e le metanalisi, pur essendo definiti studi secondari, possiedono più valore dal punto di vista della significatività degli studi primari, e vengono perciò fatte rientrare per il loro peso scientifico negli studi primari.

Studi primari

STUDIO CLINICO

Lo studio primario per eccellenza è rappresentato dallo *studio clinico* o *clinical trial* (sinonimi: *therapeutic trial*, *intervention trial*): si tratta di uno studio prospettico, in cui i dati vengono raccolti nel tempo, dopo aver stabilito lo scopo e il disegno sperimentale dello studio; l'argomento per eccellenza degli studi clinici è di tipo terapeutico. Lo studio prospettico si contrappone allo studio retrospettivo, che è invece caratterizzato dalla raccolta dei risultati dopo che un determinato evento si è verificato.

Uno studio clinico può essere uno studio controllato o uno studio controllato e randomizzato. In entrambi i casi i soggetti arruolati nello studio vengono reclutati e assegnati in modo casuale (studio controllato randomizzato) o meno (studio controllato non randomizzato) a 2 o più gruppi: il gruppo 1 riceve la terapia da testare, mentre il gruppo 2 (gruppo di controllo) riceve o un placebo (trattamento biologicamente inerte) o nessun trattamento (gruppo di controllo non trattato) o il tipo di trattamento considerato lo standard di riferimento per quella determinata patologia al momento della realizzazione dello studio. I due gruppi per poter essere confrontabili devono possedere caratteristiche quanto più possibile simili. Entrambi i gruppi vengono sottoposti a follow-up per un periodo di tempo determinato, al termine del quale vengono valutati i risultati del trattamento. Un tipo particolare di studio randomizzato è il cosiddetto *cross over* in cui i due gruppi di pazienti assumono entrambi i trattamenti, ma in successione inversa: un gruppo prima il trattamento A e poi il trattamento B, mentre l'altro prima il B e poi l'A; il momento in cui i due

gruppi cambiano trattamento è detto *cross over* (incrocio). Per realizzare questo tipo di studio è indispensabile conoscere bene le caratteristiche dei trattamenti somministrati, in modo da lasciare un tempo sufficiente di assenza di terapia (*wash out*) tale da permettere che gli effetti del primo trattamento si esauriscano prima che venga iniziato il secondo.

All'inizio di ogni studio clinico deve essere definito l'obiettivo dello studio e vanno stabiliti i parametri utilizzati per misurare l'obiettivo: i cosiddetti *endpoint* o *outcome* o misure di efficacia del trattamento. Gli *outcome* possono appartenere a tre categorie:

- clinici: mortalità, eventi morbosi, misure clinico-metaboliche (es. pressione, glicemia, ecc.);
- economici: diretti (durata ospedalizzazione, ecc.) o indiretti (giornate lavorative perse);
- umanistici: qualità di vita, presenza di sintomi, ecc.

L'ideale per un trial clinico sarebbe la misurazione di almeno un *outcome* per ogni categoria, ma in genere vengono utilizzati outcome clinici, e solo recentemente alcuni *outcomes* umanistici.

In uno studio clinico deve essere definito l'*endpoint* primario, che è quello che fornisce l'informazione più convincente e clinicamente rilevante sull'obiettivo primario dello studio, ed è quello su cui viene pianificata la numerosità del campione in studio, e gli *endpoints* secondari, che rappresentano misure aggiuntive associate all'obiettivo primario o misure di efficacia su obiettivi secondari. Si possono poi distinguere gli *endpoints* clinici maggiori o clinicamente rilevanti e gli *endpoints* surrogati: i primi sono quelli che forniscono una misura diretta dell'effetto clinico del trattamento, es. mortalità o morbilità per un certo evento, i secondi invece forniscono una misura indiretta dell'effetto del trattamento nelle situazioni in cui la misurazione diretta dell'effetto clinico non è fattibile o risulta sconveniente, es. test di laboratorio. Il problema è che nel corso degli anni si è sviluppato un ricorso smodato agli *endpoints* surrogati, che presentano indiscutibili vantaggi quali la minore durata del follow-up e la riduzione dei costi del trial. Essi tuttavia non sempre rappresentano un dato rilevante dal punto di vista clinico e talora evidenziano risultati discordanti rispetto a quelli ottenuti misurando l'*endpoint* clinicamente rilevante. Esiste un sistema di validazione degli *endpoints* surrogati, proposto dalle *User's Guides to Biomedical Literature*, che definisce per gli *endpoints* surrogati tre livelli:

- Livello 1: esiste una associazione stretta, indipendente e consistente tra *endpoint* surrogato e *endpoint* clinicamente rilevante;
- Livello 2: esistono evidenze sperimentali, condotte con altre classi di farmaci, che il miglioramento dell'*endpoint* surrogato ha determinato il miglioramento di quello clinicamente rilevante;

- Livello 3: esistono evidenze sperimentali condotte con farmaci della stessa classe, che il miglioramento dell'*endpoint* surrogato ha determinato il miglioramento di quello clinicamente rilevante.

Uno studio randomizzato può essere:

- in *aperto (open label)*: sia i partecipanti che i medici sono a conoscenza del tipo di trattamento che viene utilizzato per ogni soggetto; è possibile però mettere in atto una valutazione in cieco degli *outcome*;
- in *singolo cieco (single blind)*: solo i pazienti non sanno a quale gruppo appartengono;
- in *doppio cieco (double blind)*: né i pazienti né gli sperimentatori sanno a quale gruppo appartiene un paziente;
- in *triplo cieco (triple blind)*: si realizza quando coloro che raccolgono e analizzano i dati raccolti (*data collectors* e *data analysts*) sono diversi dagli sperimentatori e anch'essi sono ignari del gruppo a cui appartengono i soggetti arruolati.

Lo scopo del *blinding*, cioè del realizzare uno studio in cieco, è la riduzione dei *bias* cosiddetti di performance (differenza sistematica dell'assistenza erogata ai due gruppi di pazienti, conoscendo quale trattamento è stato assegnato) e di deteazione (condizionamento nella valutazione dei risultati sulla base della conoscenza del trattamento somministrato al paziente), questi ultimi soprattutto negli studi con *outcome* soggettivi (ad esempio dolore o qualità della vita).

Nonostante l'ideale per uno studio clinico sia quello di mantenere *blinded* tutti i soggetti coinvolti nello studio, non sempre questo è realizzabile. Se lo studio confronta l'effetto di due farmaci somministrati con modalità/via di somministrazione differente, per mantenere il *blinding* è necessario prevedere il doppio placebo (*double dummy technique*). Negli studi invece in cui il confronto è realizzato tra due trattamenti non mascherabili, es. terapia medica e chirurgica, o radio e chemioterapia, il *blinding* non è praticamente realizzabile.

A questo punto risulta indispensabile un breve cenno alla definizione di *bias*, traducibile in italiano con il termine distorsione o viziatura: il *bias* definisce l'errore metodologico e sistematico presente in uno studio. Tale errore si ripercuote sui risultati dello studio, determinando uno scarto tra i risultati ottenuti e quelli che si sarebbero ottenuti in assenza di *bias*. Tra le molte possibili classificazioni dei *bias* ricordiamo le tre categorie più importanti:

- *bias* di selezione: si verificano quando il campione indagato è stato scelto o assemblato in modo errato;

- *bias* di misurazione: si verificano quando i metodi di misurazione non sono ben tarati, o sono applicati in modo diverso tra i pazienti studiati;
- *bias* di confondimento, questi ultimi si verificano quando un fattore estraneo viene ad associarsi casualmente o meno sia all'esposizione-trattamento, sia all'esito.

La *randomizzazione* o assegnazione casuale dei pazienti al gruppo sperimentale o di controllo è il sistema per eccellenza per eliminare da uno studio clinico i *bias* di selezione; vi sono casi però in cui uno studio clinico non può essere randomizzato, per motivi pratici o etici (es. un trial che confronta il decorso del neonato dopo la nascita in ospedale o a casa, ecc.). Fondamentale per una corretta randomizzazione è il *concealment of allocation*, cioè l'occultamento dell'assegnazione: l'ideale a questo scopo è che il processo di randomizzazione sia realizzato da qualcuno che non sia responsabile del reclutamento dei pazienti, in modo da prevenire in tutti i modi la preconcoscenza e l'influenza esterna sul gruppo di assegnazione dei pazienti. I metodi ritenuti adeguati di randomizzazione sono: gli schemi di randomizzazione centralizzati o gestiti ad esempio dalla farmacia, che distribuisce i farmaci; l'utilizzazione di contenitori di farmaci assolutamente identici, numerati o codificati, somministrati sequenzialmente; i sistemi computerizzati in cui la sequenza di assegnazione è contenuta in un file inaccessibile (*locked*). Sono invece da evitare i sistemi di randomizzazione basati sulla assegnazione sistematica, sull'estrazione a sorte tipo "testa o croce", sul numero di cartella clinica, sulla data di nascita; questi ultimi metodi presentano un alto rischio di *bias* di selezione, in quanto l'assegnazione non è completamente occultata e può essere manipolata, si parla allora di *trial* "quasi randomizzato", non essendo protetto completamente nei confronti della possibile influenza dei selezionatori. Tipi particolari di randomizzazione sono rappresentati dalla randomizzazione a blocchi e dalla randomizzazione stratificata. La randomizzazione a blocchi ha lo scopo di bilanciare l'asimmetria quantitativa dei pazienti assegnati ai due gruppi; il blocco rappresenta una sequenza della lista di randomizzazione che contiene lo stesso numero di pazienti da assegnare ai trattati e ai non trattati. La randomizzazione stratificata è un metodo di randomizzazione finalizzato all'inserimento nei due gruppi di confronto dello stesso numero di partecipanti con una caratteristica che si pensa potrebbe influenzare la prognosi o la risposta all'intervento studiato. Gli strati sono definiti come sottogruppi della popolazione che condividono una determinata caratteristica, per es. gruppi di età o condizione socio-economica, e la randomizzazione stratificata si realizza mediante gruppi di randomizzazione separati per ogni strato.

Una classificazione importante riguarda gli studi clinici concernenti la sperimentazione di nuovi farmaci. Tali studi vengono normalmente divisi in tre fasi, le quali includono generalmente un numero crescente di soggetti. Gli studi di Fase 1 sono condotti su un numero limitato di soggetti, per lo più volontari sani, e sono finalizzati alla valutazione della sicurezza di un prodotto e dei suoi effetti collaterali. Durante questa fase vengono studiate anche le diverse dosi del farmaco, e viene stabilito il limite massimo di dosaggio per l'uomo. Il superamento di questa fase è presupposto essenziale al passaggio agli studi di Fase 2. Gli studi di Fase 2 hanno lo scopo di evidenziare come il farmaco in studio agisce sulle persone malate. Sono condotti per lo più su un numero limitato di pazienti, è presente un gruppo di controllo a cui è somministrato un placebo o un farmaco già in uso, e hanno lo scopo di dimostrare la reale efficacia del farmaco studiato. Il superamento di questa fase è presupposto essenziale al passaggio agli studi di Fase 3. Gli Studi di Fase 3 sono tipicamente studi clinici controllati e randomizzati, effettuati su un numero ampio di pazienti e con una durata superiore a quella degli Studi di Fase 2. I risultati degli Studi di fase 3 sono determinanti per la registrazione dell'utilizzo di un farmaco e delle sue indicazioni. Gli Studi di Fase 3 B sono quelli realizzati dopo l'approvazione del farmaco ma prima della sua commercializzazione. Infine vi sono gli Studi di Fase 4, che sono quelli realizzati dopo la messa in commercio del farmaco in questione, e sono spesso sponsorizzati da una Casa Farmaceutica, con lo scopo di conoscere meglio l'efficacia del nuovo farmaco nei confronti di altri farmaci già in commercio, o di conoscerne gli effetti in una particolare categoria di pazienti non studiati in precedenza.

Attualmente gli autori di studi clinici hanno a disposizione un documento, scaricabile gratuitamente da Internet anche in lingua italiana (www.consort-statement.org), messo a punto per la prima volta nel 1996 da un gruppo di clinici ed esperti statistici con il consenso delle più importanti associazioni di Editori scientifici a livello mondiale: il documento *Consort (CONSolidated Standards of Reporting Trials)*. Il *Consort* ha la finalità di fornire indicazioni per il miglioramento della stesura e della realizzazione di uno studio clinico randomizzato, indipendentemente dal contenuto scientifico e dalla leggibilità del lavoro, aspetti che non vengono presi in considerazione nel documento. Il documento *Consort* è costituito da un diagramma di flusso, che descrive le fasi di uno studio randomizzato: reclutamento, assegnazione dell'intervento, follow-up e analisi dei dati, e da una checklist delle varie voci da includere nella stesura di uno studio randomizzato divise nei vari capitoli (introduzione, metodi, risultati, discussione). I membri del comitato *Consort* si riuniscono regolarmente per garantire l'aggiornamento del documento, che rappresenta quindi il

frutto di un processo dinamico in continua evoluzione¹. L'ultima riunione del gruppo *Consort* si è svolta nel 1997, e si attende attualmente la versione aggiornata del *Consort*.

STUDIO DI COORTE

Lo studio di coorte è uno studio di tipo osservazionale (non sperimentale) analitico per lo più prospettico. Consiste nel prendere in considerazione uno o più gruppi di soggetti (coorte) con in comune una o più caratteristiche (per lo più esposizione ad un determinato fattore di rischio), e nel valutare all'interno di questa coorte lo sviluppo o meno nel tempo di una condizione in rapporto a uno o più fattori di esposizione. Lo studio di coorte può essere anche uno studio di prognosi: ad esempio un gruppo di soggetti a cui è stata diagnosticata una malattia in fase precoce o è risultato positivo un certo test viene sottoposto a follow-up per valutare il decorso o l'incidenza (nuovi casi per anno) di una nuova malattia. È un tipo di studio piuttosto costoso e lungo, può richiedere un follow-up di decenni, non è adatto allo studio di malattie con lunga latenza né di malattie rare per la difficoltà di reperimento di un numero di casi sufficiente, i risultati possono essere inficiati dai casi persi al follow-up.

L'esempio più famoso di studio di coorte è rappresentato dalla serie di tre lavori pubblicati sul *BMJ*, rispettivamente nel '64, '76 e '94, realizzati da Hill, Doll e Peto in Gran Bretagna. Gli Autori hanno studiato un gruppo di 40000 medici inglesi, suddivisi in quattro coorti sulla base dell'abitudine al fumo (non fumatori e numero di sigarette fumate al giorno), valutando a distanza di 10, 20 e 40 anni la mortalità e l'incidenza di tumore al polmone²⁻⁴. Questo studio ha permesso di dimostrare il legame tra fumo e malattia con un livello di evidenza estremamente elevato, quale può essere ottenuto con uno studio di coorte ben condotto.

STUDIO CASO-CONTROLLO

Lo *studio caso-controllo* è uno studio di tipo osservazionale analitico retrospettivo. Consiste nel confrontare un gruppo di soggetti affetti da una determinata malattia o condizione con un gruppo di controllo senza malattia o con un'altra malattia. Viene quindi ricercato il rapporto tra la malattia o la condizione in esame e l'esposizione ad un determinato fattore di rischio, sospetto di entrare in gioco nel causare la malattia. In pratica si misura l'esposizione pregressa effettuata separatamente da due gruppi diversi di soggetti, denominati appunto casi e controlli. Storicamente il primo studio caso-controllo è stato realizzato nel 1926 sull'associazione tra tumore maligno della mammella e fertilità. Questo tipo di studio possiede un basso valore di evidenza, ma può rappresentare l'unica opzione per studiare condizioni patologiche rare, dato che i casi ven-

gono raccolti in modo retrospettivo. Risponde prevalentemente a quesiti di tipo eziologico. I limiti principali sono rappresentati dalla suscettibilità a fattori di distorsione, in particolare *bias* di selezione (scelta inadeguata dei casi), specie nei confronti del gruppo di controllo, e i cosiddetti *recall bias*, determinati dalla differenza di informazione ricevuta dai casi rispetto ai controlli: i casi tendono infatti ad attribuire ad avvenute esposizioni l'insorgenza della loro patologia, per cui ricordano meglio o più intensamente l'avvenuta esposizione, e dalle difficoltà che possono essere incontrate nella selezione di un appropriato gruppo di confronto.

Lo studio *Interphone* è uno studio internazionale, che ha coinvolto 13 nazioni, sulla possibile correlazione tra utilizzo dei telefoni cellulari e rischio di tumori al capo. Lo studio è stato coordinato dall'Agenzia Internazionale per la ricerca sul cancro, patrocinato dall'Oms e finanziato dall'Ue. Lo studio rappresenta il lavoro di ricerca più esauriente mai svolto su questo tema. Si tratta di uno studio caso-controllo che ha valutato complessivamente circa 6500 pazienti affetti da uno dei seguenti 4 tipi di tumore: meningioma, glioma, neurinoma del nervo acustico o tumore della parotide. Nello studio, l'uso pregresso del cellulare da parte dei pazienti è stato messo a confronto con l'utilizzo da parte di un numero pressoché uguale di controlli del tutto simili ai pazienti oncologici per età, sesso e altri fattori, ad eccezione della malattia. Il dato che interessava rilevare era se i pazienti oncologici avessero fatto un uso più intensivo del cellulare rispetto ai controlli, e in caso positivo, se l'uso del cellulare potesse essere interpretato come un'indicazione di aumento del rischio. I risultati provvisori più importanti dello studio, in corso di pubblicazione, sono stati: (i) se si effettua una valutazione complessiva dei casi, non si evidenzia alcun aumento del rischio. (ii) Non si rileva alcun aumento del rischio per periodi di utilizzo inferiori a 10 anni. (iii) Anche per periodi di utilizzo superiori a 10 anni non si possono riconoscere indicazioni di un aumento del rischio né dei tumori del nervo acustico (neurinomi del nervo acustico) né dei tumori del tessuto cerebrale (gliomi). Tuttavia, a causa del numero relativamente piccolo di casi che presentavano un utilizzo a lungo termine, l'attendibilità statistica dei risultati appare al momento insufficiente^{5,6}.

STUDIO TRASVERSALE

Lo studio trasversale o di prevalenza (*cross sectional study*) studia la relazione esistente tra una malattia o un'altra determinata condizione e altre variabili presenti in una determinata popolazione in un determinato momento. Non richiede che i soggetti vengano seguiti nel tempo né retrospettivamente né prospettivamente, ma si tratta della fotografia istantanea di un gruppo di persone esaminate in un determinato momento; è

quindi possibile stimare la prevalenza della malattia in quella popolazione ma non l'incidenza. Dato che esposizione e malattia sono rilevati simultaneamente non è possibile stabilire la sequenza temporale tra esposizione e malattia, cioè se l'esposizione è stata precedente all'insorgenza della malattia, elemento necessario per trarre ogni tipo di inferenza causale. Pertanto è preferibile utilizzare gli studi *cross-sectional* per investigare la relazione tra caratteristiche personali che non cambiano nel tempo, come le caratteristiche genetiche e l'insorgenza di una data malattia. Questo tipo di studio non è quindi adatto, ad esempio, ad investigare fattori di rischio ambientali, ma può fornire valide indicazioni per altri tipi di indagini, come studi sui test diagnostici o sugli screening.

Studi secondari

REVISIONE

Le revisioni si distinguono in tradizionali e sistematiche.

Le revisioni tradizionali o narrative (*overview*) consistono nella raccolta di un certo numero di articoli su uno specifico argomento di cui si esegue una sintesi. L'esempio classico è la raccolta bibliografica per una tesi.

Le revisioni sistematiche (*systematic review*) sono sintesi sistematiche delle conoscenze, basate sulla valutazione critica comparata di tutta la letteratura disponibile su un determinato argomento.

Il termine sistematico si riferisce al fatto che la revisione deve essere pianificata come un vero e proprio studio, seguendo un processo strutturato ben definito che prevede:

- la formulazione di un quesito specifico (obiettivo della revisione);
- la ricerca esaustiva e oggettiva di tutti gli studi disponibili (tenere presente che su Medline sono identificabili solo il 30-80% di tutti gli studi randomizzati controllati), che dovrebbe comprendere la ricerca di studi parzialmente pubblicati (es. abstract presentati a convegni) o non pubblicati (es. tesi di laurea, rapporti di ricerca, ecc.);
- un processo di selezione e di stesura di riassunti non viziati;
- l'analisi critica delle informazioni e la sintesi delle stesse.

Le revisioni sistematiche basate sia su studi pubblicati che non pubblicati non dovrebbero risentire dei *bias* di pubblicazione. Per *bias* di pubblicazione si intende la distorsione esistente in letteratura legata al fatto che tendono ad essere pubblicati più facilmente gli studi che riportano risultati positivi. Il fatto che siano più frequentemente pubblicati gli studi che dimostrano l'efficacia di un trattamento piuttosto che quelli che ne dimostrano l'inefficacia, porta a una enfattizzazione dei risultati positivi.

La ricerca quindi degli studi non pubblicati, sebbene complessa e difficilmente esaustiva, dovrebbe essere alla base di una corretta revisione sistematica.

METANALISI

La metanalisi consiste in una revisione sistematica in cui i dati provenienti dai diversi studi, scelti per la somiglianza del loro disegno, vengono combinati quantitativamente utilizzando specifiche tecniche statistiche, al fine di generare un unico dato conclusivo relativo ad uno specifico quesito clinico. I termini metanalisi e revisione sistematica vengono talvolta usati come sinonimi, mentre in realtà molte revisioni sistematiche non procedono allo stadio di metanalisi in quanto la diversità dei disegni di studio o la natura dei dati non lo consente. Questa combinazione quantitativa non avviene sommando i risultati dei singoli studi come se fossero un unico esperimento, ma viene realizzata salvaguardando l'individualità delle stime ottenute nel singolo studio e dando a queste stime un peso proporzionale alla numerosità del campione che rappresentano.

La combinazione dei dati diminuisce l'imprecisione dei risultati dei singoli studi. Il limite di molte metanalisi è rappresentato dall'eterogeneità, cioè dall'esistenza di importanti differenze tra gli studi considerati. Si può trattare di eterogeneità statistica, quando metodi statistici differenti sono stati usati nei vari studi, oppure di eterogeneità clinica, quando le differenze riguardano il tipo di paziente, di trattamento o di esito. L'eterogeneità può rendere inaffidabile o inappropriata la combinazione quantitativa dei dati ricavati dai diversi studi.

Esiste un'organizzazione internazionale, la *Cochrane Collaboration*, il cui scopo è di aiutare le persone a prendere le decisioni più opportune sulla base di una corretta informazione, attraverso la preparazione e l'accessibilità di revisioni sistematiche aggiornate e di qualità elevata su un ampio insieme di interventi sanitari. La banca dati *Cochrane* mette a disposizione i seguenti database (www.cochrane.org):

- CDSR (*Cochrane Database of Systemic Reviews*): contiene il testo integrale delle revisioni sistematiche (*Cochrane Reviews*), aggiornate trimestralmente, ed i protocolli per le revisioni in corso di preparazione;
- CCTR (*Cochrane Controlled Trials Register Ovid*): è una raccolta bibliografica di studi clinici controllati e/o randomizzati completati, con aggiornamento trimestrale;
- DARE (*Database of Abstracts of Reviews of Effects*): contiene gli abstracts di revisioni di efficacia prodotti dagli esperti del *National Health Services for Reviews and Dissemination di York* (UK).

Dal sito *Cochrane* è possibile inoltre scaricare un glossario dei termini utilizzati nella ricerca clinica: Green S, Higgins J, editors. Glossary. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.5 [updated May 2005]. <http://www.cochrane.org/resources/handbook/>

Minor/invited papers

INVITED PAPERS

Gli *invited papers* sono articoli che vengono nella maggior parte dei casi scritti su richiesta dell'Editor della rivista da esperti in specifici campi clinici o di ricerca. Si può trattare di Editoriali, esposizione dell'esperienza personale o dei pareri di un "esperto" autorevole (*opinion leader*) su un determinato argomento-*relevant topic*, o di *review*, intesa come sintesi critica delle più recenti acquisizioni relative ad uno specifico argomento. Tipi particolari di *invited paper*, non sempre redatti su invito dell'Editor, sono i *position paper* o le *guidelines*, che rappresentano spesso il frutto del lavoro compiuto da un gruppo di esperti su incarico di una Società Scientifica Medica, con lo scopo di esporre una specifica posizione o definire linee guida su uno specifico argomento scientifico.

MINOR PAPERS

LETTERE ALL'EDITOR

Rappresentano un breve commento di un esperto su un articolo già pubblicato dalla rivista. Il commento dovrebbe essere obiettivo, costruttivo ed educativo.

PRELIMINARY REPORT O TECHNICAL NOTES

Si tratta di articoli brevi con una struttura però analoga a quella degli *original article*. Riportano i risultati preliminari su casistiche ridotte o l'esposizione di specifiche procedure tecniche.

CASO CLINICO (CASE REPORT)

Consiste nella descrizione di un caso clinico o di una serie di casi clinici (*case series*) di solito di patologie rare, in cui gli autori dovrebbero riportare, oltre alle osservazioni personali sul caso considerato, anche una revisione bibliografica dell'argomento trattato. Va tenuto presente che il valore scientifico di una singola osservazione aneddotica è basso in quanto non consente una quantificazione della probabilità che quell'evento si ripeta in una serie di pazienti. Ne consegue che un *case report* rappresenta un articolo che sarà difficilmente citato nella bibliografia di altri articoli, per cui una rivista che pubblica molti *case report*

tenderà ad ottenere un *impact factor* più basso rispetto ad una rivista che non pubblica *case report* o ne pubblica pochi. Questa è la ragione per cui numerose riviste scientifiche non accettano più di pubblicare *case report*.

The hierarchy of evidence

(qualità della documentazione di efficacia)

Il peso relativo dei diversi tipi di lavori scientifici nel prendere decisioni relativamente agli interventi clinici li pone nel seguente ordine⁷:

- a. Revisioni sistematiche e metanalisi
- b. RCT di dimensioni adeguate con *endpoints* “forti” (misure di efficacia), cioè solidi e inequivocabili, clinicamente rilevanti
- c. RCT di piccole dimensioni con *endpoints* rilevanti
- d. RCT con endpoints deboli
- e. CT non randomizzati
- f. Studi di coorte
- g. Studi caso-controllo
- h. Casi clinici
- i. Pareri di esperti autorevoli o di comitati di esperti

Per quanto riguarda invece l’adeguatezza del tipo di studio scientifico in rapporto con l’argomento trattato la suddivisione è la seguente⁸:

Argomento	Quesito	Tipo di studio
Terapia	Efficacia di una terapia medica o di un trattamento chirurgico	Studio clin. randomizzato
Diagnosi	Validità e ripetibilità di un esame diagnostico	Studio trasversale (confronto con gold standard)
Prognosi	Evoluzione di un soggetto affetto da una determinata malattia ad un certo stadio	Studio di coorte Studio caso controllo
Eziologia	Relazione tra l’esposizione ad una determinata sostanza e lo sviluppo di una certa malattia	Studio di coorte Studio caso-controllo (in base alla rarità della malattia)
Screening	Validità di un test applicato a un gran numero di persone	Studio trasversale

Come scrivere correttamente un lavoro scientifico

Nel momento in cui si decide di scrivere un lavoro scientifico, la prima domanda da porsi è senza dubbio se si ha qualcosa in mente per cui valga veramente la pena di intraprendere uno sforzo così grande, quale è quello di portare a termine la stesura di un articolo scientifico. Per scoprirlo è indispensabile eseguire una ricerca della letteratura scientifica aggiornata sull'argomento. Se la risposta è affermativa è necessario essere in grado di rispondere a qualche altra domanda prima di iniziare il lavoro: che cosa si vuole dimostrare nel lavoro? Si è in grado di aggiungere qualcosa di nuovo a quello che è già stato scritto in passato sull'argomento? Che tipo di studio si ha intenzione di realizzare? A chi è indirizzato e a quale rivista si pensa di proporlo per la pubblicazione?

L'ICMJE (*International Committee of Medical Journal Editors*), noto anche come gruppo di Vancouver dalla sede dove per la prima volta si è riunito, ha messo a punto un documento *Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication*, che ha lo scopo di fornire un aiuto sia agli autori che agli editori nel creare e distribuire articoli scientifici di qualità. Il documento è diviso in due parti: una prima parte riporta le raccomandazioni relative ai principi etici che devono essere alla base del processo di valutazione e pubblicazione dei manoscritti di argomento biomedico e dei rapporti tra gli autori, i revisori e gli editori. La seconda parte è invece più tecnica, e riguarda le indicazioni pratiche per la preparazione e la stesura degli articoli scientifici. Il documento è disponibile gratuitamente su Internet (www.ICMJE.org), l'ultima versione disponibile è aggiornata a ottobre 2008.

La struttura di un articolo sia esso osservazionale o sperimentale deve essere diviso nelle seguenti quattro sezioni, secondo la struttura cosiddetta IMRAD: Introduzione, Materiali e metodi, Risultati e Discussione. Sono possibili varianti terminologiche a seconda della rivista, ad esempio *Background* invece di *Introduction*, o *Subjects and Methods* invece di *Materials and Methods*, ma la suddivisione non cambia. Tale struttura non è in realtà una semplice arbitraria divisione in capitoli, ma riflette il processo con cui viene elaborato un articolo scientifico. Analizziamo brevemente le varie parti di cui un articolo si compone.

Titolo: è estremamente importante; basta pensare che negli anni 80 veniva stimato che per ogni persona che leggeva un articolo scientifico altri 500 ne leggevano solo il titolo. Tale proporzione si è sicuramente modificata con la diffusione della disponibilità di letteratura scientifica su Internet, ma il titolo di un articolo mantiene tutta la sua importanza: pur mantenendosi conciso, esso dovrebbe contenere in una semplice frase che cosa è stato studiato e con quali risultati. Per lo più è conveniente

utilizzare titoli descrittivi, che contengano esclusivamente il messaggio fondamentale del risultato ottenuto, risultando così di maggior impatto per il lettore. È possibile formulare titoli interrogativi, ma è preferibile un titolo assertivo a meno che non si tratti di una review, che consideri molte possibili risposte ad una domanda. Da evitare l'utilizzo di sigle o acronimi nel titolo, ad eccezione delle sigle universalmente riconosciute (es. OSAS).

Abstract: convenzionale o strutturato a seconda delle norme editoriali della rivista a cui il manoscritto viene inviato, deve rispecchiare rigorosamente il contenuto dell'articolo e deve quindi contenere: gli scopi del lavoro, gli aspetti principali dei metodi, i risultati significativi (se possibile i dati numerici che possiedono una significatività statistica), le principali conclusioni. Non deve invece contenere: informazioni non riportate nell'articolo, un eccessivo uso di abbreviazioni (meglio se nessuna) e riferimenti bibliografici. Accanto all'abstract le riviste scientifiche richiedono la presenza delle parole chiave (*key words*), in numero variabile da 3 ad un massimo di 10; se possibile devono essere tratte dalla terminologia dell'indice usato dalla *United States National Library of Medicine* per indicizzare gli articoli sull'Index Medicus e su MEDLINE, cioè il *Medical Subjects Headings* (MeSH).

Introduzione: ha lo scopo di esporre i motivi che hanno portato alla realizzazione dello studio, per cui dovrebbe rispondere ai seguenti quesiti: perché la ricerca eseguita è importante? Quali ricerche sono state realizzate in precedenza su quello stesso argomento e con quali limiti? Qual'è l'ipotesi che si va a cercare di dimostrare e con quale impatto sulla pratica clinica? Bisogna ricordare che conviene evitare di dilungarsi eccessivamente nell'introduzione, piuttosto conviene trasferire nella discussione il materiale che potrebbe far superare i limiti di lunghezza consigliati per l'introduzione (1-2 pagine).

Materiali e Metodi: lo scopo di questo capitolo è di riportare tutte le informazioni che possano consentire in linea teorica la ripetizione dello studio su un campione di soggetti analogo, e quindi di confermare o negare i risultati ottenuti. Il capitolo deve contenere esclusivamente le informazioni disponibili all'inizio dello studio, mentre tutte le informazioni ottenute successivamente saranno riportate nel capitolo dei risultati. La definizione della condizione o malattia in studio, i criteri di selezione dei pazienti (criteri di inclusione e criteri di esclusione) e per quali ragioni tali criteri sono stati applicati e le procedure utilizzate per la selezione devono essere riportati accuratamente. In questa sede devono essere dichiarati l'approvazione da parte del Comitato Etico, l'ottenimento del consenso informato da parte dei pazienti studiati e, nel caso si tratti di uno studio randomizzato, il sistema di randomizzazione applicato. La metodica

utilizzata deve essere descritta brevemente con gli opportuni riferimenti bibliografici se si tratta di una metodica già utilizzata, oppure dettagliatamente se si tratta di una metodica di nuova introduzione. Vanno riportate le misure di *outcome*, sia principali che secondarie. È indispensabile ricordarsi di specificare i dati tecnici relativi alle strumentazioni utilizzate (nome della ditta, città e stato). Infine deve essere descritta in modo preciso la metodologia statistica utilizzata⁹.

Risultati: una regola fondamentale è l'applicazione della corrispondenza biunivoca tra Metodi e Risultati: tutti i risultati ottenuti devono trovare nel capitolo Metodi la descrizione di come sono stati ottenuti e viceversa tutti i metodi descritti devono trovare l'esposizione dei relativi risultati ottenuti in questo capitolo. I risultati devono essere presentati in sequenza logica, partendo dai più generali per passare ai più specifici, evitando qualsiasi considerazione che ne valorizzi o svaluti il significato. I risultati numerici vanno sempre riportati, non ci si può accontentare dei dati da essi derivati, ad esempio delle percentuali. È consigliabile l'utilizzazione di tabelle o grafici, per riportare i dati dei risultati, evitando di ripetere nel testo tutti i dati numerici già esposti nelle tabelle. Tuttavia non bisogna eccedere con le tabelle e le figure: il loro numero va limitato allo stretto indispensabile a supportare il testo. Le tabelle e le figure devono essere tutte numerate, citate nel testo e dotate di una didascalia che ne esponga in modo chiaro il contenuto, secondo la regola di base che tabelle e figure dovrebbero essere comprensibili anche per chi le osserva senza aver letto l'articolo di cui fanno parte integrante. In questo capitolo vanno riportati i dati sulla significatività statistica o meno dei dati ottenuti, precisando il valore numerico del p sia quando esso è significativo che non.

Discussione: rappresenta la parte più "libera" dell'intero articolo, il suo scopo è di fornire un'interpretazione e un commento ai risultati ottenuti, evitando l'espressione di opinioni personali o generali sul tema dello studio. Essa è costituita da un insieme di osservazioni basate sul riesame critico dello studio realizzato, che deve tener conto dei seguenti punti: risposta al quesito o all'ipotesi posta all'inizio della ricerca ed esposto nell'introduzione; nel caso si tratti di uno studio sperimentale, esposizione delle possibili teorie per spiegare i dati rilevati e confronto dei risultati ottenuti con i dati presenti in letteratura; analisi dei limiti dello studio (es. campione troppo piccolo, selezione non casuale, presenza di *bias*) e del grado di generalizzabilità dei risultati a una popolazione più grande, con eventuale riferimento a possibili sviluppi in studi futuri¹⁰.

Alcune riviste richiedono un capitolo separato per le conclusioni, mentre altre preferiscono che siano inserite al termine della discussione. In ogni caso una conclusione dovrebbe essere breve e concisa, dovrebbe esprimere un giudizio ragionato basato esclusivamente sui risultati dello stu-

dio realizzato, ma non dovrebbe rappresentarne un riassunto. Da evitare tassativamente di riportare nelle conclusioni affermazioni non supportate dai risultati dello studio, o speculazioni basate su teorie personali, non provate né documentate.

Bibliografia: la scelta e la precisione della citazione delle voci bibliografiche nel rispetto completo delle norme editoriali della rivista a cui si intende sottoporre il manoscritto per pubblicazione è fondamentale per la sua valutazione. Tutte le voci bibliografiche citate devono essere richiamate nel testo dell'articolo. Nella compilazione della bibliografia è consigliabile limitarsi a riportare le voci bibliografiche più rilevanti e recenti, evitando articoli non reperibili su MEDLINE/Pubmed, ad eccezione per la citazione di libri o altri volumi, es. atti congressuali. Da evitare di riportare voci bibliografiche "di seconda mano", cioè riportate da altri articoli, senza aver verificato la correttezza della citazione.

Qualità delle pubblicazioni scientifiche

La qualità delle pubblicazioni scientifiche è un concetto che non presenta al momento ancora una definizione precisa e universale; la sua valutazione può basarsi su un duplice approccio: qualitativo o quantitativo, espresso in termini numerici di impatto scientifico.

A partire dagli anni '70 le pubblicazioni scientifiche sono state oggetto di studio da parte della bibliometria, disciplina basata sullo studio quantitativo delle pubblicazioni, che ha lo scopo di analizzare i modelli di distribuzione delle pubblicazioni e esplorarne l'impatto nell'ambito delle comunità scientifiche. Tuttavia gli indicatori bibliometrici rischiano di non essere in grado di valutare la reale qualità delle singole pubblicazioni, per cui l'ideale sarebbe arrivare ad una metodica di valutazione basata sulla combinazione di analisi quantitativa e qualitativa.

Attualmente il metodo più accreditato per garantire la qualità delle pubblicazioni scientifiche, difendendo il lettore da pubblicazioni non rispondenti ai canoni di scientificità o realizzate esclusivamente con finalità di vantaggi personali, è rappresentato dal filtro di qualità del ciclo editoriale, in particolare attraverso la procedura del giudizio dei "pari valutatori" o *peer review*. Un lavoro scientifico inviato ad una rivista può essere rifiutato direttamente dall'editore, nel caso non corrisponda alle finalità del periodico o presenti una evidente bassa qualità, oppure può essere inviato alla revisione critica da parte di esperti dell'argomento, in genere due. Nel caso in cui i due esperti esprimano un parere discordante, l'articolo viene inviato ad un terzo esperto. Generalmente l'identità del revisore viene mantenuta segreta, mentre l'autore dell'articolo è noto al revisore; alcune

riviste, circa il 40%, utilizza la procedura *double blind* in cui esperti e autori sono entrambi anonimi, mentre al contrario nel sistema *open peer review*, utilizzato da un numero limitato di periodici, la revisione viene attuata *on line* e sono conosciuti reciprocamente sia gli autori che i revisori. Quest'ultimo sistema è adottato al fine di limitare i rischi di abusi o atti di malafede, sia da parte degli autori (produzione di più articoli con i risultati della stessa ricerca, omissione di autori che hanno collaborato alla ricerca...) che da parte dei revisori (plagio di pubblicazioni, deliberato ritardo di valutazione di articoli potenzialmente competitivi, discriminazione mirata...). Esiste attualmente quasi completo accordo sul fatto che il sistema della *peer review* debba essere mantenuto come sistema di garanzia di qualità delle pubblicazioni scientifiche¹¹. Proposte alternative a tale procedura sono legate alla diffusione delle pubblicazioni *on line*, in particolare è stato proposto ed inizia ad essere utilizzato l'*open peer commentary*, un sistema in cui si passa da una valutazione eseguita da un ristretto numero di esperti, quale la *peer review*, alla pubblicazione su Internet sia del lavoro da recensire che dei giudizi espressi dagli esperti, consentendo di aumentare il numero dei potenziali valutatori e di rendere trasparente ed aperto il giudizio dei revisori. Un altro tipo di approccio già sperimentato prevede dapprima la messa in linea di un articolo da parte di un autore per sollecitare gli esperti ad esprimere un parere su di esso e solo successivamente, dopo le dovute modificazioni, l'invio dell'articolo alla rivista prescelta.

Per quanto riguarda invece la valutazione di tipo quantitativo, attualmente l'indicatore bibliometrico maggiormente utilizzato è rappresentato dall'*Impact Factor* (IF).

L'IF è calcolato e pubblicato annualmente dall'Institute of Scientific Information (ISI), che è un'impresa privata statunitense fondata da Eugene Garfield, nata con l'idea di creare un repertorio delle pubblicazioni scientifiche, che comprendesse l'elenco dei lavori citati da ogni lavoro considerato. Sulla base di questa idea è stato creato un archivio elettronico delle citazioni denominato *Science Citation Index* (SCI), che comprende un numero selezionato di riviste pressoché esclusivamente in lingua inglese. La selezione delle riviste è svolta a discrezione dell'ISI basandosi sulle caratteristiche principali di una rivista scientifica, in particolare i criteri di valutazione richiedono: la puntualità nella pubblicazione dei fascicoli, l'applicazione di una procedura di *peer review* per la valutazione editoriale degli articoli; la presenza di un *abstract* in inglese, per le poche riviste non pubblicate interamente in lingua inglese; l'internazionalità degli autori; l'interesse per il contenuto scientifico in relazione anche alla attuale copertura della specifica categoria tematica o alla trattazione di argomenti emergenti.

L'elenco delle riviste viene aggiornato frequentemente, determinando continuamente l'ingresso di nuove riviste e la fuoriuscita di altre, con una tendenza generale di aumento del numero complessivo delle riviste.

L'IF è basato sull'ipotesi che il numero di citazioni di un articolo sia direttamente proporzionale alla sua diffusione all'interno della comunità scientifica, e rappresenti quindi il suo valore scientifico. Esso viene calcolato con una formula numerica che definisce l'IF di una rivista X nell'anno Y come il rapporto tra il numero di citazioni di articoli pubblicati sulla rivista X rilevate nei due anni precedenti a Y, diviso per il numero totale di articoli pubblicati dalla rivista X in quegli stessi due anni. Sulla base di questo calcolo è compilata annualmente una classificazione delle riviste con un relativo punteggio.

Come ogni strumento, l'IF ha dei limiti: la quasi totale esclusione delle riviste pubblicate in lingua non inglese, con una notevole prevalenza delle riviste anglo-americane tra quelle incluse nella banca dati dell'ISI, il che rende particolarmente difficoltosa una comparazione internazionale; in secondo luogo l'ISI è una struttura privata con finalità di lucro, per cui nelle sue decisioni non possono essere esclusi interessi economici; infine il sistema dell'IF rischia di condizionare le modalità di compilazione delle bibliografie dei lavori scientifici, ad esempio la promozione delle autocitazioni, cioè la rivista che cita se stessa. Ma il limite più importante è insito nell'ipotesi stessa che la qualità scientifica di un articolo possa essere valutata mediante un indicatore puramente quantitativo¹².

Tuttavia, nonostante questi limiti l'IF resta lo strumento a disposizione più utilizzato e "obiettivo" per la valutazione della produzione scientifica di un autore a livello accademico. Le riviste di argomento otorinolaringoiatrico dotate di IF secondo l'elenco aggiornato al 2008 sono 31, tutte pubblicate in lingua inglese e tra esse le 3 ad IF più alto sono rappresentate da Head & Neck, il Journal of the Association for Research in Otolaryngology e Audiology & Neuro-Otology, le prime due riviste sono statunitensi mentre la terza è europea, pubblicata in Svizzera. Al momento nessuna rivista italiana di argomento otorinolaringoiatrico risulta impattata, ma la rivista italiana Acta Otorhinolaryngologica Italica, organo ufficiale della Società Italiana di Otorinolaringoiatria, è stata accettata nella banca dati dell'ISI, ed è attualmente in fase di valutazione per il raggiungimento dell'IF.

Per concludere è doveroso ricordare che le tecniche bibliometriche si stanno evolvendo rapidamente nel tempo in parallelo con la progressiva sempre maggiore diffusione del Web. Sono sorti quindi nuovi database e nuovi strumenti che permettono la ricerca di citazioni, e parallelamente nuovi indici bibliometrici alternativi all'IF. Tra questi ricordiamo:

l'Eigenfactor (EI) (www.eigenfactor.org) rappresenta un nuovo modello matematico per calcolare l'impatto delle riviste scientifiche che tiene conto dei diversi standard di citazione tra una disciplina e l'altra;

l'Indice di Hirsch o H-index, proposto nel 2005 da Jorge E. Hirsh, finalizzato alla quantificazione della prolificità e dell'impatto del lavoro dei singoli autori basandosi sul numero delle loro pubblicazioni ed il numero di citazioni ricevute. La sua importanza è legata al fatto che è in grado di quantificare la reale influenza di un autore sulla comunità scientifica, prescindendo dai singoli articoli pubblicati di grande successo, ma anche dai molti articoli di scarso interesse, come può avvenire invece utilizzando l'IF;

Publish or Perish (www.harzing.com/resources.htm#/pop.htm) è un software scaricabile gratuitamente che recupera e analizza le citazioni scientifiche e consente di calcolare una serie di indici bibliometrici di uno specifico autore, tra cui il Normalized Individual H-index, basato sulla normalizzazione del numero di citazioni per ogni articolo diviso per il numero degli autori dell'articolo stesso, calcolando poi l'H-index sul conteggio normalizzato delle citazioni;

SCImago Journal & Country Rank (www.scimagojr.com) è un database, lanciato nel dicembre del 2008, accessibile gratuitamente, che permette di generare statistiche sulle citazioni degli articoli pubblicati nelle riviste peer-reviewed. Il database è in grado di generare statistiche per paese, confrontando il numero degli articoli pubblicati e le citazioni differenziate per paese di provenienza.

Bibliografia

1. Moher D, Cook DJ, Eastwood S, Olkin I, Renne D, Stroup DF. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-4.
2. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. *BMJ* 1964;248(i): 1399-1410, 1460-67.
3. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *BMJ* 1976;273(ii):1525-36.
4. Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994;309: 901-11.
5. Hours M, Bernard M, Montestrucq L, Arslan M, Bergeret A, Deltour I, Cardis E. Cell Phones and risk of brain and acoustic nerve tumours: the French INTERPHONE case-control study. *Rev Epidemiol Sante Publique* 2007;55:321-32.
6. INTERPHONE Study Group. Brain tumour risk in relation to mobile telephone use: results of the INTERPHONE international case-control study. *Int J Epidemiol* 2010;39:675-94.
7. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' guides to the medical literature. IX. A method for grading health care recommendations. *JAMA* 1995;274:1800-4.
8. Greenhalgh T. How to read a paper: getting your bearings (deciding what the paper is about). *BMJ* 1997;315:243-6.
9. Sardanelli F, Di Leo G. Come si scrive un lavoro scientifico radiologico. In: Sardanelli F, Di Leo G. editors. *Biostatistica in Radiologia*. Milan: Springer: 2008. p.177.220.
10. Gustavii B. How to write and illustrate a scientific paper. New York. Cambridge University Press: 2003.
11. Rosenfeld RM. How to review journal manuscripts. *Otolaryngol Head Neck Surg* 2010;142:472-86.
12. Dracos A, Cognetti G. La letteratura scientifica. Indicatori bibliometrici e bibliografici quali criteri integrativi per la valutazione obiettiva di un'attività di ricerca. *Ann Ist Super Sanità* 1995;31:381-90.

ANALISI CRITICA DI UN ARTICOLO SCIENTIFICO
ASPETTI GENERALI
S. Palma, M. Piemonte

Lo scopo di un articolo scientifico di carattere clinico ben fatto è dare un'informazione nuova (diagnostica, terapeutica, eziologia, prognostica) che possa essere utilizzata da altri clinici nella loro professione.

In questo capitolo si vuole sottolineare alcuni aspetti importanti della metodologia clinica che devono essere presenti nelle diverse sezioni dell'articolo.

L'articolo scientifico deve dunque possedere una rilevanza clinica: avere, ad esempio, la capacità di evidenziare l'effetto di un trattamento quando questo effettivamente esiste (validità interna, scientifico-metodologica) e consentire l'applicazione delle conclusioni dello studio su popolazioni reali di pazienti (validità esterna, generabilità).

In base alla rilevanza clinica lo studio illustrato nell'articolo può raggiungere un determinato livello di evidenza e un conseguente grado di raccomandazione. In linea generale il livello di evidenza è espresso con un numero accompagnato da una lettera minuscola dell'alfabeto, quanto più piccolo è il numero con la lettera "a", maggiore è il livello di evidenza. Il grado di raccomandazione è espresso con una lettera maiuscola (A: fortemente raccomandato, C: poco o nulla raccomandato).

Per la tabella riassuntiva dei livelli di evidenza e di raccomandazioni si rimanda al capitolo 2.

Introduzione

Come già esposto nel precedente capitolo 3, in questa parte dell'articolo sono esposti i motivi dello studio e una succinta sintesi dell'argomento già riportato in Letteratura.

Lo scopo dello studio è l'elemento chiave per dare risalto all'articolo, esso esprime l'ipotesi che gli Autori intendono verificare. Si completa con le conclusioni: se queste sono rilevanti, dedotte da una corretta analisi dei risultati, lo scopo annunciato assume particolare rilevanza e interesse. Nell'introduzione è bene specificare il tipo e il disegno dello studio, ovvero se è uno studio di tipo prospettico o retrospettivo, se è un trial randomizzato oppure uno studio osservazionale, e così via.

Metodi

In questa parte dell'articolo vengono indicati i pazienti arruolati nello studio e le tecniche di misura degli *outcomes*. È in questa parte che è descritta la reale conduzione dello studio, ed è pertanto in questa fase che si possono compiere una buona parte di errori metodologici.

La conduzione dello studio prevede la scelta degli endpoints, la selezione dei pazienti con o senza randomizzazione a seconda del tipo di studio (applicando criteri di inclusione e di esclusione), determinare la numerosità del campione, la programmazione del follow-up.

1) Endpoints. Devono essere chiari, ben descritti, inequivocabili, validati oggettivamente (es. scale del dolore). Devono essere solidi, clinicamente rilevanti. È preferibile ad esempio valutare una sopravvivenza a lungo termine piuttosto che a breve. Altra caratteristica da considerare è la facile misurabilità: è preferibile usare metodiche diffuse e standardizzate piuttosto che nuovissime, ma poco fruibili. Se si ricorre all'uso di questionari è meglio utilizzare quelli brevi piuttosto che lunghi, che sono sì più esauritivi, ma poco attendibili perché riducono la compliance del paziente.

2) Numerosità del campione. Sulla selezione dei pazienti, soprattutto negli studi randomizzati, si rimanda agli altri capitoli. Qui vogliamo sottolineare l'importanza del numero di soggetti da arruolare nello studio perché l'effetto stimato sia il più possibile coincidente con quello reale. L'effetto del trattamento infatti non sempre coincide con quello reale perché si può avere l'interferenza di fattori confondenti, rappresentati da due tipi di errori: *errori sistematici* ed *errori casuali*. Gli errori sistematici devono essere evitati attraverso appropriati accorgimenti in fase di pianificazione, conduzione ed analisi dei risultati dello studio e non risentono della numerosità del campione. Gli errori sistematici di fatto costituiscono i *biases* (viziature) dello studio. Nella tab. I sono indicati i principali tipi di bias.

L'adeguata numerosità del campione consente di ottenere la *significatività statistica* dello studio, derivata dall'applicazione di un test statistico ai risultati. Il valore di questo test indica se l'effetto osservato è imputabile almeno al 95% alla sperimentazione e non deriva dalla casualità.

Gli errori casuali derivati da una inadeguata numerosità sono di due tipi, definiti come *errore tipo I* ed *errore tipo II*.

Supponiamo di sperimentare l'efficacia di un farmaco "X" (es. antibiotico o immunostimolante) nel trattamento delle faringo-tonsilliti acute valutando il numero di episodi acuti per anno. Ipotizziamo di selezionare 10

Tipo di bias	Descrizione
di selezione	Si verifica quando il campione è stato scelto e assemblato in modo errato
di misurazione	Si verifica se i metodi di misurazione non sono ben tarati o validi, oppure se sono imprecisi o diversi tra i pazienti studiati
di confondimento	Si verifica quando è presente un fattore esterno non controllato, che è associato, anche se in modo non causale, sia alla esposizione-trattamento, sia all'esito
di pubblicazione	È dovuto alla tendenza a pubblicare preferenzialmente risultati positivi e a non pubblicare (tacitare) i risultati negativi

Tab. I: tipi principali di bias

pazienti per il gruppo trattato e 10 per il gruppo placebo-controllo dalla nostra popolazione.

Se dividessimo tutta la popolazione in due gruppi, uno trattato e l'altro di controllo-placebo, potremo ottenere un andamento dei risultati come illustrato dalle due curve gaussiane indicate nella parte A e B della fig. 1, dove per semplicità abbiamo indicato sulle ascisse solo il numero di episodi faringo-tonsillitici per anno. Dall'analisi statistica delle differenze delle medie delle due sottopopolazioni otteniamo ad esempio che il trattamento è effettivamente efficace. Se per il nostro studio estraiamo 10 persone da ciascuna sottopopolazione (indicati con i cerchietti), avremo una rappresentazione limitata della distribuzione dei risultati (parte C della fig. 1). Analizzando i dati ottenuti solo dai soggetti selezionati potremo ottenere un valore del nostro test statistico che non ci permette di affermare che il trattamento è efficace. Su questo concetto si basano gli errori tipo I e II.

L'errore di tipo I deriva dal concetto di "ipotesi nulla". Con questa espressione si vuole intendere che il presupposto iniziale dello studio è che il trattamento non è efficace; solitamente l'interesse degli sperimentatori è rifiutare l'ipotesi nulla affermando così che il trattamento è efficace. L'errore di tipo I è rifiutare l'ipotesi nulla, quando questa in realtà è vera, cioè affermare che il trattamento è efficace quando nella realtà non lo è. Per convenzione si accetta un rischio di commettere questo tipo di errore (probabilità α) non superiore al 5% (sinteticamente indicata con la seguente espressione: $p \leq 0.05$). In altre parole se si eseguisse l'esperimento per 100 volte con la stessa numerosità del campione per 5 volte corriamo il rischio di esprimere un giudizio errato sul trattamento.

L'errore di tipo II si basa invece sull'ipotesi che il trattamento sia inefficace (non rifiutare l'ipotesi nulla). In questo caso l'errore è derivato dall'accettare l'ipotesi di inefficacia, quando in realtà è falsa (= trattamento efficace). In altre parole si stabilisce la probabilità di trarre una conclusione

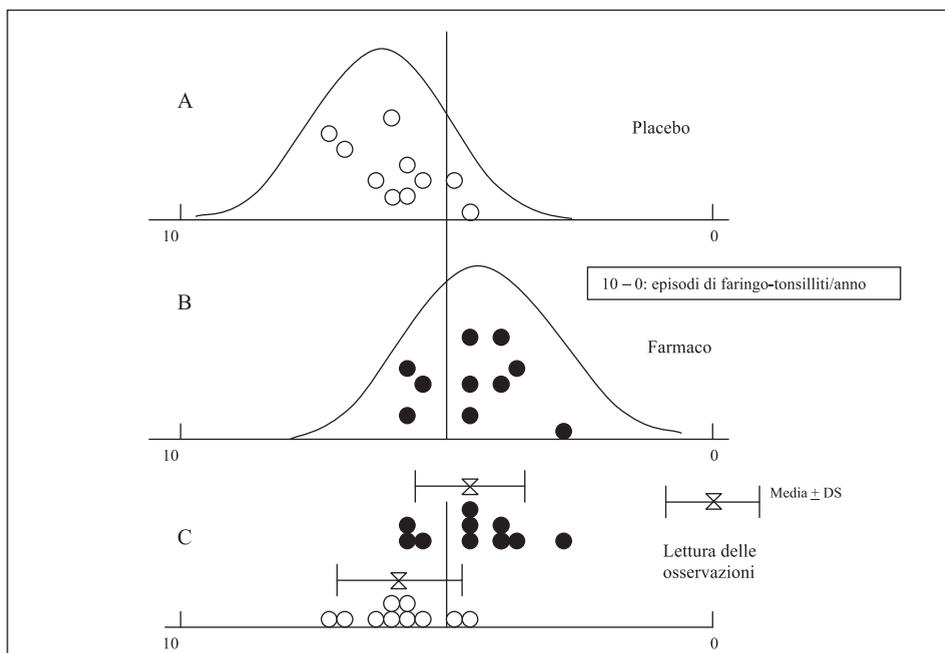


Fig. 1: rappresentazione grafica dello studio ipotizzato (vedi testo); i cerchi (pieni per il gr. trattato, vuoti per il gr. controllo) rappresentano i soggetti estratti per lo studio, le curve rappresentano la reale distribuzione dell'effetto del trattamento nelle due popolazioni

falsamente negativa, cioè di non riportare un effetto quando invece esiste. Solitamente per convenzione si accetta la probabilità, indicata con β , di commettere questo errore non superiore al 20%.

La probabilità di rilevare un "vero positivo", cioè di evidenziare una differenza statisticamente significativa è espressa dal complemento di β , ovvero $1 - \beta$. Questo valore indica la *potenza dello studio* e si ritiene statisticamente significativo quando non è inferiore a 0,80. Ciò indica la probabilità dell'80% di uno studio di evidenziare un effetto che nella realtà esiste. Nella tab. II sono indicati i tipi di conclusioni possibili in un test statistico di un'ipotesi.

Conclusioni tratte dalle osservazioni	Situazione reale	
	Il trattamento è efficace	Il trattamento è inefficace
Il trattamento è efficace	Vero positivo Conclusione corretta ($1 - \beta$)	Falso positivo Errore di tipo I (α)
Il trattamento è inefficace	Falso negativo Errore di tipo II (β)	Vero negativo Conclusione corretta ($1 - \alpha$)

Tab. II: tipi di conclusioni possibili in un test statistico di un'ipotesi

Su questi due parametri α e β gli statistici stabiliscono la numerosità del campione. Se l'effetto del trattamento è consistente (ad esempio risoluzione di un edema laringeo dopo 10 minuti dalla somministrazione endovenosa di un cortisonico) il numero di soggetti da arruolare può essere contenuto. Se invece l'effetto di un trattamento è meno evidente (esempio farmaco antivertiginoso) allora la numerosità deve essere più cospicua.

In sintesi un lavoro scientifico ben fatto riporta un adeguato numero di pazienti costituente il campione in esame tale che i risultati abbiano dal punto di vista statistico una certa significatività, ovvero che α sia inferiore al 5% e $1 - \beta$ sia uguale o superiore all' 80%.

3) Sistema "P.I.C.O."

P.I.C.O. è un acronimo inglese che sta per: Patient/population, Intervention, Comparison/comorbidity, Outcome.

Questo semplice schema è molto utile da seguire quando si redige la parte di materiali e metodi di un articolo scientifico.

I pazienti o la popolazione sono quelli arruolati nello studio. Di essi è importante definire l'età, il sesso e soprattutto la malattia che rappresenta l'obiettivo principale dello studio. In particolare se lo studio prevede 2 gruppi, uno sottoposto ad uno specifico trattamento, l'altro costituente il gruppo controllo o placebo, la malattia deve essere distribuita in maniera uniforme, anche come gravità tra i due gruppi. Questo dovrebbe essere garantito da una adeguata randomizzazione. I pazienti vengono arruolati in funzione dei criteri di inclusione ed esclusione. Solitamente i criteri di esclusione prevedono la gravidanza o l'allattamento (a meno di studi *ad hoc* per questa categoria di donne) e, nel caso di impiego di farmaci, intolleranza accertata verso i farmaci impiegati nello studio.

I criteri di inclusione dovrebbero essere il meno restrittivi possibile, affinché i dati ottenuti possano essere più facilmente applicati ad una popolazione generale e rendere quindi i risultati dello studio clinicamente molto rilevanti. La descrizione dell'intervento prevede l'azione diretta dello sperimentatore sul campione dei pazienti in esame. L'intervento può essere anche semplicemente di tipo osservazionale se l'obiettivo è valutare un fattore di rischio. Negli studi di terapia vanno descritti accuratamente tipi e dosaggi di farmaci impiegati, nei casi in cui l'atto terapeutico è chirurgico va descritta succintamente la tecnica chirurgica e specificate le tecnologie adottate. Il confronto deve essere fatto o con terapie mediche già standardizzate o con un placebo somministrato con le stesse modalità del farmaco in studio e comunque non identificabile. Negli studi di chirurgia il confronto può essere fatto o con altre tecniche chirurgiche o con altre terapie non chirurgiche.

Nei casi di studi diagnostici il test diagnostico in esame deve essere ben illustrato tale da poter essere ripetibile e riproducibile e deve essere indicato il test diagnostico di confronto che costituisce al momento il gold standard per quella patologia oggetto dello studio (tale aspetto sarà sviluppato nell'apposito capitolo).

Uno studio ben condotto prevede anche la specificazione delle comorbidità dei pazienti oggetti dello studio. Questo è particolarmente importante se nello studio sono previsti due gruppi a confronto, perché una ineguale distribuzione delle comorbidità potrebbe falsare i risultati ottenuti. Infatti i processi di guarigione possono essere più lenti nei pazienti con più comorbidità, inoltre tali pazienti sono soggetti più facilmente a complicanze.

Gli outcomes sono strettamente legati all'obiettivo dello studio. Gli outcomes vanno suddivisi in principale e secondari. Il principale outcome deve avere forte rilevanza clinica (es. mortalità) e costituisce indubbiamente l'asse portante dello studio. Gli outcomes secondari, ma non per questo non importanti, emergono dall'elaborazione dei dati raccolti e danno ulteriori informazioni che possono essere comunque utili dal punto di vista clinico (es. effetti indesiderati osservati durante la sperimentazione di una terapia farmacologica).

4) Programmazione del follow-up. La durata del follow-up deve ovviamente essere proporzionata all'effetto del trattamento che si deve misurare. È importante quando si imposta uno studio randomizzato in doppio cieco che l'intensità e la numerosità degli eventi non siano troppo piccoli. È indispensabile che il tipo e la frequenza dei controlli siano uguali nei due gruppi di trattamento.

Altro elemento importante per ridurre il numero di uscita dallo studio è l'adeguata informazione dei pazienti, che devono essere stimolati a continuare il trattamento indipendentemente dalla sensazione soggettiva di efficacia.

Risultati

Per l'analisi quantitativa dei risultati si rimanda agli specifici capitoli.

Qui vogliamo enfatizzare l'importanza dell'approccio qualitativo delle analisi di uno studio controllato randomizzato.

Importanza dell'approccio "Intention-to-treat".

Quando si conduce uno studio clinico controllato randomizzato si considera un numero di soggetti che per le caratteristiche patologiche possono essere arruolati nello studio. Il numero però viene ridimensionato in

base ai criteri di inclusione e di esclusione, ottenendo il gruppo di soggetti eleggibili. Una quota di questi non accetta di partecipare allo studio (i così detti *refuser*), il numero si riduce così a quelli che effettivamente iniziano lo studio (soggetti randomizzati, assegnati ad uno dei due gruppi dello studio).

I soggetti che iniziano lo studio, assegnati a caso ad uno dei due trattamenti, sono definiti “intention to treat (ITT)” (letteralmente: intenzione a trattare). Alcuni di questi soggetti, pur essendo entrati in uno dei due bracci dello studio, potrebbero non iniziare affatto il trattamento assegnato. Alcuni soggetti durante il trattamento e durante il follow-up abbandonano lo studio (*drop-out*). Il numero di soggetti che completano tutto lo studio (*compliant*) è solitamente inferiore a quello dei soggetti che iniziano o che hanno mostrato intenzione di iniziare il trattamento. Alcuni soggetti possono lasciare lo studio anche per decisione dello sperimentatore (*discontinuer*), esempio per aggravamento della malattia o per comparsa di altra malattia o per grave evento avverso.

In uno studio clinico controllato randomizzato si possono quindi verificare 4 situazioni:

1. deviazione dal trattamento A: soggetti che vengono persi o che passano nell'altro trattamento B (*drop in*).
2. Nessuna deviazione dal trattamento A: i soggetti completano il trattamento A a loro assegnato.
3. Deviazione dal trattamento B: altra quota di soggetti persi o che passano nell'altro trattamento A.
4. Nessuna deviazione dal trattamento B: i soggetti completano il trattamento B a loro assegnato.

A questo punto si possono definire 3 principali approcci di analisi dei dati:

- a. *ITT-analysis* (analisi in base all'intenzione di trattare): si valuta i soggetti con il relativo trattamento assegnato: (1) e (2) versus (3) e (4).
- b. *Adherers-only (AO) analysis* (analisi in base all'adesione al trattamento): si valuta solo i soggetti che hanno completato il trattamento a loro assegnato: (2) versus (4).
- c. *As-treated (AT) analysis* (analisi in base al trattamento completato): si valuta i soggetti che hanno terminato uno dei due trattamenti (il trattamento effettivamente completato indipendentemente da quello iniziato): (1) e (4) versus (2) e (3).

L'analisi ITT è quella da preferire, in quanto mantiene la comparabilità delle caratteristiche dei partecipanti dei due gruppi. Inoltre questo tipo di analisi produce risultati che rispecchiano più verosimilmente le con-

dizioni reali; infatti nella pratica clinica comune si verifica facilmente che alcuni pazienti non seguano o abbandonino il trattamento prescritto.

Un'analisi AO è forse di più immediata comprensione per il clinico, abituato a prescrivere una terapia ed a verificarne l'effetto mediante controlli. Con questo tipo di analisi si valutano i risultati dei soggetti che hanno completato il loro trattamento assegnato, ma vengono lasciati fuori quelli che hanno deviato dal protocollo. Questo però può far sì che i due gruppi alla fine del trattamento non siano più omogenei e differiscano per caratteristiche rispetto alla situazione iniziale.

Un'analisi AT è in un certo senso l'inverso dell'analisi ITT in quanto sono presi in considerazione non i soggetti a cui è stato assegnato un trattamento, ma i soggetti che hanno ricevuto un trattamento alla fine dello studio. Non viene preso in considerazione il problema di quale trattamento sia durato più a lungo nei soggetti che hanno effettuato il passaggio da un braccio all'altro dello studio.

Se uno studio è di tipo randomizzato in doppio cieco e la perdita di soggetti è contenuta (entro il 10%) con una distribuzione più o meno uniforme tra i due gruppi, è verosimile che la perdita sia pure "randomizzata" e questo non dovrebbe modificare le caratteristiche generali dei due gruppi, pertanto i diversi approcci di analisi non dovrebbero dare risultati molto discostanti tra loro e dare una indicazione abbastanza uniforme sull'effetto del trattamento che si vuole studiare.

Se invece lo studio non è in doppio cieco, ma aperto, come può succedere ad esempio nel confronto tra terapia farmacologica ed altro tipo di terapia (fisioterapica, chirurgica), la fuoriuscita dal trattamento assegnato per abbandono o per passaggio all'altro trattamento, ritenuto dal soggetto più efficace, può portare ad una alterazione dei due gruppi di trattamento sia in termini numerici che qualitativi per sbilanciamento delle caratteristiche di base.

Questo problema viene ovviato con l'analisi ITT. Il problema di questo tipo di analisi è come utilizzare i soggetti drop-out, ovvero come inserire i loro dati nell'analisi globale. Non vi è consenso unanime su quale metodica utilizzare per manipolare i dati dei soggetti persi e nessuna metodica può considerarsi ottimale. Una metodica semplice è quella di utilizzare come risultato il valore disponibile al momento dell'assegnazione del trattamento (misura dell'outcome pre-trattamento), oppure se sono già state eseguite delle valutazioni in corso del trattamento si utilizza il dato dell'ultimo rilievo ("*carry-forward imputation*"). Un altro metodo è quello di utilizzare la media dei valori dei risultati ottenuti al termine del follow-up dei pazienti allocati nello stesso braccio dello studio che hanno terminato il trattamento ("*mean imputation*").

Per comprendere meglio quanto finora esposto proviamo a fare un esempio.

Gruppo trattamento				Gruppo placebo		
N. pz	prima	dopo		N. pz.	prima	dopo
1	7	4		1	6	5
2	6	3		2	6	6 4,125
3	8	5		3	7	3
4	9	9	2,875	4	8	4
5	6	2		5	9	4
6	7	1		6	6	6 4,125
7	7	4		7	7	5
8	8	2		8	8	6
9	6	6	2,875	9	8	4
10	9	2		10	6	2

Tab. III: ipotetico studio immunostimolante versus placebo; sono riportati il numero di episodi faringo-tonsillitici/anno in due anni di osservazione: il primo senza trattamento il secondo con trattamento (farmaco o placebo).

Riferendoci all'esempio ipotetico del secondo paragrafo, immaginiamo di aver raccolto i dati relativi al numero di episodi faringo-tonsillitici per anno dopo somministrazione di un farmaco, esempio immunostimolante, versus placebo, come riportato nella tab. III.

I 10 pazienti selezionati per ogni gruppo mostrano un indice pre-trattamento superiore a 5 (criterio di inclusione). Supponiamo che vengano persi 2 pazienti per ciascun gruppo, rispettivamente i numeri 4 e 9 del gruppo trattamento e numeri 2 e 6 del gruppo placebo.

Nella nostra esposizione dei dati di questo studio immaginario ci limitiamo al dato grezzo relativo alla differenza delle medie, senza inoltrarci a indicare test statistici, come andrebbe effettuato nella realtà, in quanto esula dalle finalità di questo testo l'approfondimento delle analisi statistiche. Se si eseguisse un'AO *analysis* non prenderemo in considerazione i soggetti persi, e le medie verrebbero effettuate su 8 soggetti per gruppo. Se invece volessimo eseguire una *ITT analysis* allora considereremo tutti i 20 soggetti, utilizzando per i soggetti persi o l'ultimo dato a disposizione (equivalente al valore pre-trattamento: *carry forward imputation*) oppure il valore medio degli altri 8 valori post-trattamento a disposizione (*mean imputation*). Nella tab. IV sono riportate le differenze delle medie pre- e post-trattamento intragruppo e intergruppo in funzione dei diversi modi di analisi.

Come si osserva nella tab. IV la differenza trattamento – placebo è a favore del primo trattamento, però l'entità varia a seconda del tipo di analisi. In questo tipo di studio la situazione più vicina alla realtà è probabilmente l'analisi tipo *ITT carry-forward*, in quanto è abbastanza verosimile che i soggetti abbandonino il trattamento perché non ne hanno beneficio, mentre i soggetti che ne traggono beneficio sono più

Gruppo	Tipo di analisi		
	<i>AO analysis</i>	<i>ITT carry-forward</i>	<i>ITT mean</i>
Trattamento	4,375	3,5	4,425
Placebo	3,25	2,6	2,975
Trat. – Placebo	1,125	0,9	1,45

Tab. IV: diverse analisi delle differenze delle medie dei risultati riportati in tab. III

propensi a proseguire il trattamento e quindi giungeranno al termine dello studio.

Vi sono altre metodiche di manipolazione dei dati dei soggetti persi nella ITT analysis: assumere il numero dei soggetti persi come risultato peggiore oppure come risultato migliore, assegnare i casi persi come migliori ad un trattamento e come peggiori all'altro (situazione "estrema" a favore di uno o dell'altro trattamento). Nella tab. V riportiamo un esempio ipotetico esplicativo di questi concetti. Come si può osservare i dati riportati nell'ultima colonna possono variare anche di molto dando un risultato a favore di uno o dell'altro trattamento a seconda della metodica usata, soprattutto se la percentuale di pazienti persi è molto alta.

In ogni caso l'analisi ITT non può risolvere i problemi e le viziature che possono emergere se il disegno dello studio non è stato ben progettato e i sistemi di misura non sono efficaci e validi.

In alcuni casi è giusto non considerare nell'analisi (deviare dall'approccio ITT) i soggetti reclutati nella randomizzazione che non abbiano eseguito il trattamento assegnato. Infatti può accadere che alcuni soggetti vengano reclutati quando i dati a disposizione per soddisfare i criteri di inclusione non sono ancora disponibili (esempio disponibilità di un dato di laboratorio, ma necessità di iniziare il trattamento quanto prima) e solo dopo l'avvio del trattamento o addirittura al suo compimento si scopre che tali soggetti non necessitavano del trattamento e quindi non soddisfavano i criteri di inclusione. Questa situazione è facilmente verificabile nella realtà clinica, in quanto spesso i clinici devono iniziare un trattamento prima di avere a disposizione tutti i dati diagnostici (esempio comune: iniziare un trattamento antibiotico basandosi sui dati clinici prima di avere l'esame microbiologico che può risultare negativo). Anche in questo caso ricorrere comunque ad una ITT analysis può essere un procedimento corretto per quanto sopra esposto.

Una situazione più chiara in cui non ricorrere all'approccio ITT è corretto è quando il trial prevede un trattamento chirurgico. I soggetti che non vengono sottoposti al trattamento chirurgico in studio, anche se arruolati, non possono fornire risultati in qualche modo assimilabili a quelli dei soggetti operati.

	Decongestione chirurgica (n = 100)	Riduzione con radiofrequenza (n = 100)	Differenza assoluta nella percentuale di ristenosi nasale
Ristenosi nasale dopo 6 mesi			
Si	34	29	-
No	40	51	-
Ignoto	26	20	-
Percentuale di ristenosi (%)			
AO analysis	46 (34/74)	36 (29/80)	10
Assunzione di risultato peggiore	60 (60/100)	49 (49/100)	11
Assunzione di risultato migliore	66 (66/100)	71 (71/100)	- 5
Situazione estrema a favore della decongestione chirurgica	66 (66/100)	49 (49/100)	11
Situazione estrema a favore della radiofrequenza	60 (60/100)	71 (71/100)	- 11

Tab. V: effetto dei vari metodi di manipolazione dei dati persi in un trial (ipotetico) di soggetti con ostruzione nasale per ipertrofia dei turbinati, sottoposti a decongestione chirurgica sottomucosa dei turbinati o a riduzione volumetrica con radiofrequenza.

Discussione e Conclusioni

Per gli aspetti della discussione si rimanda la capitolo 3.

Nelle conclusioni vengono riportate le risposte, derivate dall'analisi dei risultati, alle ipotesi dello studio formulate nell'introduzione.

Le conclusioni finali le trae il lettore, che deve considerare la generabilità dei dati offerti dallo studio, cioè se i risultati sono applicabili alla popolazione dei pazienti con quel problema e soprattutto se sono applicabili ai propri pazienti con quello stesso problema.

Bibliografia

Wright CC, Sim J. *Intention-to-treat approach to data from randomized controlled trials: A sensitivity analysis*. J Clin Epidemiol 2003; 56: 833-842.

Hollis S, Campbell F. *What is meant by intention to treat analysis? Survey of published randomised controlled trials*. BMJ 1999; 319: 670-674

**ANALISI CRITICA DELLE PROCEDURE DIAGNOSTICHE.
INDICI DI PERFORMANCE DEL TEST**

S. Palma

Nella pratica clinica la diagnosi viene ricercata attraverso l'esecuzione di diverse indagini: anamnesi, esame obiettivo, indagini strumentali radiologiche e altre di pertinenza specialistica, esami di laboratorio, esami citologici ed istologici.

Il medico dopo la valutazione clinica, spesso aspecifica e non dirimente, è obbligato ad effettuare altre indagini diagnostiche. La capacità del medico è quella di sintetizzare gli esiti dei diversi esami diagnostici eseguiti, escludendo alcune patologie e orientando la diagnosi verso altre, scegliendo infine quella che in base all'esperienza maturata e ai dati clinico-strumentali in possesso risulta la più probabile.

La condizione ideale è avere a disposizione un unico test diagnostico appropriato per ogni tipo di patologia. Attualmente il test diagnostico che dà con una certa ragionevolezza una diagnosi sicura è l'esame istologico. Tuttavia questo esame è ottenibile dopo intervento chirurgico su un pezzo operatorio e talvolta solo post mortem, oppure ricorrendo a manovre invasive per eseguire una biopsia. Vi sono comunque altri esami strumentali, propri per ciascuna Specialità, che consentono di ottenere una diagnosi quasi sempre certa o con minimi margini di errore sulla gravità della patologia.

Il test diagnostico che individua con maggior sicurezza rispetto ad altri test una determinata patologia costituisce il *gold standard* per quella patologia. A volte per ottenere una diagnosi è necessario ricorrere a più esami tra loro complementari, pertanto il gold standard diagnostico in questo caso è rappresentato da una batteria di esami.

Da diverso tempo ormai i clinici sono sempre più orientati ad utilizzare il minor numero possibile di esami diagnostici, sia per ampliamento delle conoscenze scientifiche, sia per necessità di contenimento della spesa sanitaria. L'orientamento attuale è quello quindi di utilizzare il minor numero possibile di esami diagnostici e possibilmente uno solo, cercando di ottenere con ragionevole certezza una giusta diagnosi. Da qui la continua necessità di studiare con prove di efficacia clinica test diagnostici nuovi o già in uso per validarne la loro reale utilità in campo clinico. I test diagnostici si possono suddividere in dicotomici e non dicotomici. Quelli *dicotomici* rispondono ad un criterio di positività o negatività: se il test è positivo si suppone che ci sia la malattia, se è negativo si

suppone che non vi sia la malattia. Il test dicotomico per eccellenza è l'esame istologico per neoplasia maligna. Il test può assumere diversi valori, ma quando supera un certo valore ritenuto significativo diventa positivo. Sotto a tale valore il test è ritenuto negativo. Tale valore limite è definito anche *cut off*. Un esempio tipico è rappresentato dalla misurazione della pressione arteriosa: il limite normale è ritenuto 140/90 mmHg, al di sopra di questi valori si può parlare di ipertensione sistolica e/o diastolica.

Quando invece non è possibile stabilire con ragionevole certezza un *cut off* si parla di test diagnostici *non dicotomici*. In questo caso i valori del test sono distribuiti in maniera continua da valori minimi a valori massimi, quando i valori si collocano agli estremi è ragionevole parlare rispettivamente di test negativo e test positivo, ma quando i valori si avvicinano verso quelli centrali il test rimane dubbio. Vedremo più avanti come è possibile in qualche modo cercare un *cut off* che possa orientare nella diagnosi. In ogni caso anche per i test non dicotomici è possibile stabilire arbitrariamente un *cut off* e valutare poi l'impatto clinico con la metodologia di seguito esposta.

Quando viene studiato un test diagnostico nella pratica clinica, è necessario che esso sia confrontato con il gold standard diagnostico attualmente in uso. Può comunque accadere che il nuovo test possa risultare più preciso ed efficace del gold standard finora in uso e che ovviamente andrà poi a sostituire.

Gli strumenti a disposizione per saggiare l'efficacia del test sono gli indici di performance del test, meglio conosciuti come *sensibilità*, *specificità* (che costituiscono le proprietà invarianti del test), *valore predittivo positivo* e *valore predittivo negativo* (che costituiscono le proprietà del test dipendenti dalla frequenza di malattia).

Il tasso di veri positivi (o veri negativi) esprime la percentuale di positivi (o negativi) al test dei soggetti, che effettivamente presentano la condizione morbosa (o che non la presentano) per la quale il test diagnostico è stato utilizzato.

Per comprendere meglio la sua espressione matematica consideriamo la proporzione:

$$n : N = x : 100$$

dove x esprime il valore del tasso, n (*enne piccolo*) è il numero di soggetti risultati positivi (o negativi) compreso in N (*enne grande*), che esprime la grandezza della popolazione sottoposta al test effettivamente portatrice della condizione morbosa in esame (oppure di cui ne è esente).

Pertanto più il rapporto n/N tende ad 1, più il test risulta efficace.

Il **tasso dei veri positivi (VP)** esprime la **sensibilità** del test:

$$\text{Sensibilità} = \text{VP} / \text{Totale soggetti malati} \times 100$$

	M +	M -	Tot.
Test positivo	a	b	a + b
Test negativo	c	d	c + d
Tot.	a + c	b + c	a + b + c + d

Tab. I: schema di tabella 2x2

Il **tasso dei veri negativi** (VN) esprime la **specificità** del test:

$$\text{Specificità} = \text{VN} / \text{Totale soggetti sani} \times 100.$$

Per comprendere ancora meglio questi concetti è bene utilizzare una tabella 2 x 2 (vedi anche capitolo 1).

Le righe di questa tabella (tab. I) riportano gli esiti del test, positivo nella prima, negativo nella seconda; le colonne riportano il reale stato dei soggetti esaminati, nella prima (M+) sono conteggiati i soggetti che realmente hanno la malattia, determinato con il gold standard diagnostico, nella seconda (M-) i soggetti sani (o che non hanno la malattia in esame). Il totale delle colonne riportano quindi quanti soggetti sono realmente affetti dalla malattia in esame e quanti invece non lo sono.

I totali delle righe invece riportano quanti sono risultati positivi al test e cioè ipoteticamente affetti dalla malattia e quanti invece ipoteticamente non lo sono (negativi al test).

Pertanto dall'incrocio dei dati della tabella 2 x 2 risulta che il valore indicato con "a" rappresenta il valore dei veri positivi (VP), cioè quanti soggetti effettivamente portatori della malattia sono stati identificati come tali dal test diagnostico. Il valore indicato con "d" indica il valore dei veri negativi (VN), cioè quanti soggetti effettivamente sani sono stati identificati come tali dal test.

Il valore "b" indica invece i *falsi positivi* (FP), cioè i soggetti realmente sani, ma indicati come malati dal test.

Il valore "c" indica i *falsi negativi* (FN), cioè i soggetti realmente malati, ma riconosciuti come sani dal test.

Da questi concetti si evince già che un test ideale dovrebbe avere il minor numero possibile di falsi positivi e di falsi negativi.

Per una più facile lettura della tabella 2 x 2 sostituiamo le lettere con i simboli indicati tra parentesi:

	M +	M -	Tot.
Test positivo	VP	FP	Sogg. Positivi
Test negativo	FN	VN	Sogg. Negativi
Tot.	Malati	Sani	Popolazione studiata

Tab. II: tabella 2x2 per l'analisi degli indici di performance del test

Avendo a disposizione questi dati è possibile calcolare la sensibilità, specificità, valore predittivo positivo e negativo:

1. **Sensibilità** = $VP/(VP + FN)$ = $VP/(\text{totale soggetti realmente malati})$;
2. **Specificità** = $VN/(VN + FP)$ = $VN/(\text{totale soggetti realmente sani o esenti dalla malattia in esame})$;
3. **Valore predittivo positivo** = $VP/(VP + FP)$ = $VP/(\text{totale soggetti positivi al test})$;
4. **Valore predittivo negativo** = $VN/(VN + FN)$ = $VN/(\text{totale soggetti negativi al test})$.
5. Per completezza indichiamo qui il calcolo dell'accuratezza del test, parametro peraltro poco usato: **Accuratezza** = $(VP + VN)/(VP + VN + FP + FN)$ = $\text{Soggetti correttamente diagnosticati}/\text{Totale popolazione studiata}$. Tale parametro ha poca utilità in quanto test con la stessa accuratezza possono variare ampiamente di sensibilità e specificità.

La sensibilità e la specificità sono calcolate in "verticale". Sono definite proprietà invarianti del test in quanto dipendono dalla reale efficacia del test e non dalla prevalenza di malattia e dovrebbero presentare valori stabili se il test è applicato nello stesso contesto di popolazione. Il condizionale è d'obbligo in quanto l'esito del test dipende dalla sua corretta esecuzione e interpretazione.

Il valore predittivo del test (positivo o negativo) è legato alla prevalenza della malattia nella popolazione studiata; facendo riferimento alla nostra tabella 2 x 2 la prevalenza risulta dal rapporto tra soggetti malati e soggetti studiati ($M+/ \text{popolazione studiata}$). Questo rapporto viene definito anche come **probabilità pre-test**.

Per capire meglio il valore predittivo facciamo un esempio. Supponiamo di eseguire l'endoscopia nasale per diagnosticare una rinosinusi- site cronica ad una popolazione di 20 soggetti di cui 10 sono effettivamente malati e 10 sani, potremo rilevare ad esempio 7 soggetti positivi (sensibilità del 70%) e 1 falso positivo (pari al 10% della popolazione sana studiata). Il VPP sarà di $7/8$ pari a 87,5%. Se però la popolazione cambia e i soggetti studiati sono 100 di cui 80 sono realmente sani e solo 20 sono i soggetti rinosinusitici, la prevalenza non sarà più del 50% ($10/20$) come prima, ma sarà del 20%. Il numero di falsi positivi aumenta proporzionalmente al numero di soggetti non ammalati ed essendo pari al 10%, nella seconda popolazione il numero di falsi positivi sarà di 8. Pertanto nella seconda popolazione di 100 soggetti, mantenendo l'endoscopia nasale la sua sensibilità e specificità, avremo 14 soggetti malati individuati correttamente ($14/20$), ma il valore predittivo positivo sarà del 63,6% ($14/(14+8)$), inferiore a quello

precedente dove la prevalenza della malattia era maggiore. Analogo ragionamento si può fare per il valore predittivo negativo, considerando la specificità e la percentuale di falsi negativi in funzione della prevalenza della malattia. In altre parole a parità di sensibilità e specificità la probabilità che un soggetto sia ammalato si riduce con la riduzione della prevalenza di malattia (ovvero proporzionale alla probabilità pre-test).

Pertanto in presenza di un'alta prevalenza di malattia il test diagnostico consente di confermare la patologia già sospettata, perciò la specificità dovrebbe essere alta (bassa percentuale di falsi positivi). Un sostanziale guadagno di informazione è dato dal risultato positivo del test a conferma della diagnosi. Tuttavia un risultato negativo non esclude completamente la malattia a fronte di una percentuale, anche se bassa, di falsi negativi.

Se la prevalenza di una malattia è bassa un singolo test positivo può essere non particolarmente utile nello stabilire la diagnosi in un singolo paziente, poiché la probabilità di malattia rimane bassa. Un risultato negativo aggiunge poche informazioni in quanto conferma un'alta probabilità pre-test di assenza di malattia. Ciò risulterà più chiaro quando si parlerà del rapporto di verosimiglianza.

Questo concetto comunque è da tenere presente quando si vuole applicare un test diagnostico sulla propria popolazione di riferimento. Infatti è importante sapere qual è l'incidenza e prevalenza della malattia, che si vuole testare, nella propria popolazione di riferimento. Generalmente negli studi la probabilità pre-test è più alta della reale prevalenza in quanto viene testato un campione selezionato afferente spesso in strutture di riferimento per quella malattia. Il test diventa tanto più significativo quanto più grande è il campione esaminato tale da avere una probabilità pre-test più vicina alla reale prevalenza della malattia. Quindi si richiama il lettore a porre sempre attenzione alla popolazione di riferimento dello studio e alla sede dove è stato eseguito al fine di valutare criticamente l'applicabilità del test.

Un altro parametro che può essere impiegato per valutare l'efficacia del test è l'**indice di Youden (J)** che permette di comparare i valori del test in termini di specificità e sensibilità, assumendo che questi due parametri devono essere ugualmente importanti. Questo indice si esprime nel seguente modo:

$$J = \text{sensibilità} + \text{specificità} - 1,$$

risulta chiaro che l'indice tenderà a 1 quando il test è molto efficace dando bassissimi valori di falsi positivi e di falsi negativi.

Prima di procedere oltre ed esporre i **rapporti di verosimiglianza (Likelihood Ratios: LR)**, riteniamo utile chiarire i concetti fin qui esposti con degli esempi.

ESEMPIO N.° 1

Da una casistica personale di soggetti sottoposti a tiroidectomia totale per noduli tiroidei solitari o dominanti relativa al periodo gennaio 2009 – marzo 2010, abbiamo estrapolato la seguente tabella (tab. III) che riporta l’esito dell’esame istologico postoperatorio incrociato con l’esame citologico pre-operatorio.

Nella Tab. III le righe indicano la risposta all’esame citologico del FNA sotto guida ecografia, le colonne indicano l’esito dell’esame istologico postoperatorio (la reale natura dei noduli tiroidei esaminati), le prime colonne indicano un esito istologico maligno (carcinoma papillifero, follicolare, altre neoplasie maligne tra cui carcinoma midollare e metastasi), le colonne successive verso destra indicano la natura benigna dei noduli (gozzo, adenoma, tiroidite). La colonna più a destra indica il rilievo di incidentalomi relativi a microcarcinomi papilliferi da ritenersi riscontri occasionali. A scopo esemplificativo se leggiamo la prima riga osserviamo che su 45 esami citologici (escludendo i 5 casi di incidentaloma) risultati benigni, 39 (36 + 3) sono risultati effettivamente benigni all’esame istologico definitivo, mentre 6 (3 + 3) sono risultati maligni (falsi negativi).

Nella valutazione che seguirà non sono conteggiati i soggetti con risposta citologica di “neoplasia follicolare” in quanto per definizione ancora oggi l’esame citologico non è in grado di distinguere con certa ragionevolezza tra adenoma follicolare e carcinoma follicolare. Non vengono conteggiati nemmeno gli incidentalomi, riscontri occasionali non prevedibili in fase pre-operatoria, ma che comunque cambiano la diagnosi definitiva. Pertanto la popolazione complessiva studiata è di 91 soggetti (3 incidentalomi sono compresi nel gruppo follicolare già escluso).

I soggetti M+ sono quelli con neoplasia maligna all’esame istologico che

CITOLOGICO	ISTOLOGICO						Totale
	Carcinoma papillifero	Carcinoma follicolare	Altre malignità	Patologia benigna	adenoma	incidentaloma	
Benigno	3	3	0	36	3	5	50 (35,7%)
Follicolare	4	5	0	7	25	3	44 (31,4%)
Sospetto	35	1	4	4	2		46 (32,9%)
Totale	42	9	4	47	30	8	140 (100%)

Tab III: confronto citologia/istologia dei noduli tiroidei di soggetti sottoposti a tiroidectomia totale).

	M +	M -	Tot.
Test positivo	40	6	46
Test negativo	6	39	45
Tot.	46	45	91

Tab. IV: valori di VP, VN, FP, FN estrapolati dalla tab. III

costituisce il nostro test diagnostico gold standard. M- sono gli altri soggetti senza neoplasia maligna.

I soggetti con test positivo sono quelli che all'esame citologico presentavano una risposta di sospetta neoplasia maligna o un pattern francamente maligno.

Inseriamo quindi i dati a disposizione nella nostra tabella 2 x 2

Avremo quindi i seguenti risultati:

Sensibilità = $40/46 \times 100 = 86,9\%$.

Specificità = $39/45 \times 100 = 86,7\%$.

VPP = $40/46 \times 100 = 86,9\%$.

VPN = $39/45 \times 100 = 86,7\%$.

In sintesi possiamo affermare che in base a questi dati l'esame citologico è in grado di dare una risposta corretta in almeno 8 pazienti su 10, a fronte di una probabilità pre-test molto alta del 50,5% (46/91).

In generale un test molto specifico implica la presenza di pochi falsi positivi e quindi quando è positivo è utile per confermare una malattia; un test molto sensibile prevede un basso numero di falsi negativi e quindi quando è negativo permette di escludere la malattia.

ESEMPIO N.° 2

Il secondo esempio è tratto dal seguente articolo: "A comparative study between MR sialography and salivary gland scintigraphy in the diagnosis of Sjögren Syndrome"; *H. Tonami, K. Higashi, M. Matoba, H. Yokota, I. Yamamoto, S. Sugai.* J Comput Assist Tomogr, 2001, Vol 25 (2): 262-8.

In questo studio è stata valutata l'efficacia nella diagnosi della Sindrome di Sjögren (SS) della scialo-Risonanza Magnetica (sRM) e della scintigrafia delle ghiandole salivari maggiori (ScGS).

Il gold standard diagnostico è costituito dall'esame istologico delle ghiandole salivari minori prelevate mediante biopsia labiale.

La sRM e la ScGS sono state considerate positive quando si evidenziava un quadro patologico apprezzabile anche se minimo (stadio 1 in una scala da 0 a 4 per entrambi gli esami, utilizzata dagli AA).

I pazienti reclutati sono stati 130, affetti da xerostomia con sospetto di SS, di cui 80 sono risultati effettivamente affetti dalla SS, come da esame istologico.

	M +	M -	Tot.
Test positivo	58 69	0 25	58 94
Test negativo	22 11	50 25	72 36
Tot.	80 80	50 50	130 130

Tab. V: risultati relativi allo studio riportato nell'esempio 2

I risultati dello studio possono essere riassunti nella nostra tabella 2 x 2 (tab. V) come segue (il primo numero a sinistra in grassetto è relativo alla sRM, quello a destra in corsivo alla ScGS).

Avremo quindi:

Probabilità pre-test: $80/130 = 61,5\%$

Sensibilità sRM: $58/80 \times 100 = 72,5\%$

ScGS: $69/80 \times 100 = 86,2\%$.

Specificità

sRM: $50/50 \times 100 = 100\%$

ScGS: $25/50 \times 100 = 50\%$.

VPP

sRM: $58/58 \times 100 = 100\%$

ScGS: $69/94 \times 100 = 73,4\%$.

VPN

sRM: $50/72 \times 100 = 69,4\%$

ScGS: $25/36 \times 100 = 69,4\%$.

Da notare che per la ScGS il VPP raggiunge appena il 73,4% con un probabilità pre-test del 61,5%.

In base a questi dati gli AA dell'articolo concludono che in un paziente con xerostomia e sospetta SS, il test diagnostico di scelta è la sRM. Se questo esame documenta segni della patologia la biopsia labiale può essere evitata, poiché l'elevata specificità e il VPP consentono di confermare la diagnosi.

Questo esempio illustra come le proprietà invarianti del test abbiano una efficacia limitata nella scelta diagnostica operata dal clinico di fronte ad un singolo paziente. Infatti la sensibilità è superiore per ScGS rispetto a quella della sRM. Al contrario la specificità è superiore per la sRM a fronte di una assenza di falsi positivi per questo test diagnostico. Ciò comporta un VPP del 100%.

Il likelihood ratio (LR) o rapporto di verosimiglianza e la probabilità post-test

Il problema principale per il clinico non è tanto sapere la potenzialità diagnostica di un test, ma quanto sapere qual è la probabilità di malattia nel singolo paziente adottando quel particolare test diagnostico, ovvero la probabilità di corretta diagnosi nel singolo paziente.

Box n.° 1. Odds. La parola inglese odds si può tradurre in italiano come “probabilità a favore”. È un’espressione tipica degli scommettitori inglesi (bookmakers) per quantificare la probabilità di vincere. Ad esempio se la probabilità di vittoria di una squadra o di un giocatore viene data a 4:1, equivale a dire che su una scala da 1 a 5 le probabilità di sconfitta (p) sono considerate 4 volte più alte di quelle di vittoria (1-p). La vittoria verrebbe pagata 4 volte la cifra scommessa.

Per rispondere a questo problema torna utile il likelihood ratio.

Il LR esprime il rapporto tra la percentuale dei soggetti malati positivi al test (= sensibilità) e la percentuale dei positivi al test nei soggetti sani (=falsi positivi). Questo rapporto indica quante volte è più probabile trovare il test positivo nei malati piuttosto che nei controlli.

Le formule per calcolare il LR sono le seguenti:

Rapporto di verosimiglianza positivo (LR+) =
$$[VP/(VP + FN)]/[FP/(FP + VN)] = \text{Sensibilità}/(1 - \text{Specificità})$$

Rapporto di verosimiglianza negativo (LR-) =
$$[FN/(VP + FN)]/[VN/(FP + VN)] = (1 - \text{Sensibilità})/\text{Specificità}$$

Una delle proprietà che rende interessante per il clinico il LR è che partendo da una stima della prevalenza della malattia (probabilità pre-test) con una semplice moltiplicazione è possibile conoscere la probabilità che il nostro paziente abbia la malattia (*probabilità post-test*), cioè il valore predittivo cui come clinici siamo interessati. Per ottenere questa informazione dobbiamo però trasformare la prevalenza di malattia nei cosiddetti *odds* (vedi box n. 1). Gli odds rappresentano il rapporto di due probabilità.

La formula generale per ricavare gli odds pre-test dalla prevalenza percentuale di malattia è la seguente:

$$\text{odds pre-test} = \text{prevalenza}/(100-\text{prevalenza}).$$

La moltiplicazione del LR per gli odds pre-test ci dà gli *odds post-test* cioè la probabilità di malattia nel paziente col test positivo. Avremo quindi:

$$\text{odds post-test} = LR \times (\text{odds pre-test}) = LR \times [\text{prevalenza}/(100 - \text{prevalenza})]$$

È possibile a questo punto convertire in probabilità gli odds post-test con la formula seguente:

$$\text{probabilità post-test} = \text{odds post-test} / (\text{odds post-test} + 1).$$

Avendo a disposizione anche solo la sensibilità e specificità con queste formule è possibile calcolare il LR e la probabilità post-test.

Proviamo ora ad applicare questi concetti ai nostri esempi sopra riportati.

ESEMPIO 1

Probabilità pre-test (prevalenza) = $46/91 = 50,5\%$
 $LR+ = \text{Sensibilità} / (1 - \text{Specificità}) = 0,869 / (1 - 0,867) = 6,53$
 $LR- = (1 - \text{Sensibilità}) / \text{Specificità} = (1 - 0,869) / 0,867 = 0,15$
 $\text{Odds pre-test} = \text{prevalenza} / (100 - \text{prevalenza}) = 50,5 / (100 - 50,5) = 1,02$
 $\text{Odds post-test} = LR \times (\text{odds pre-test}) = 6,53 \times 1,02 = 6,66$
 $\text{Probabilità post-test} = \text{odds post-test} / (\text{odds post-test} + 1) = 6,66 / (6,66 + 1) = 0,87 = 87\%$.
 Equivale a dire che un paziente con il test positivo, ovvero con agoaspirato di nodulo tiroideo sospetto per neoplasia maligna, ha l'87% di probabilità di avere la malattia.

ESEMPIO 2

Probabilità pre-test (prevalenza); $sRM = 80/130 = 61,5\%$
 $ScGS = 80/130 = 61,5\%$
 $LR+ = \text{Sensibilità} / (1 - \text{Specificità}); sRM = 0,725 / (1 - 0,99) = 72,5$.
() ipotizzando a scopo esemplificativo 99% anziché 100% altrimenti la divisione sarebbe impossibile*
 $LR- = (1 - \text{Sensibilità}) / \text{Specificità}; sRM = (1 - 0,725) / 0,99 = 0,28$
 $ScGS = (1 - 0,862) / 0,50 = 0,276$
 $\text{Odds pre-test} = \text{prevalenza} / (100 - \text{prevalenza});$
 $sRM = 61,5 / (100 - 61,5) = 1,6$
 $ScGS = 61,5 / (100 - 61,5) = 1,6$
 $\text{Odds post-test} = LR \times (\text{odds pre-test}) \quad sRM = 72,5 \times 1,6 = 116$
 $ScGS = 1,724 \times 1,6 = 2,76$
 $\text{Probabilità post-test} = \text{odds post-test} / (\text{odds post-test} + 1);$
 $sRM = 116 / (116 + 1) = 0,991 = 99,1\%$.
 $ScGS = 2,76 / (2,76 + 1) = 0,734 = 73,4\%$.

Questi dati indicano che un soggetto con xerostomia e sospetta SS con una scialo-RM con segni positivi di malattia ha il 99,1%, ovvero quasi la certezza, di avere la malattia; mentre con una scintigrafia delle ghiandole salivari positiva ha il 73,4% di probabilità di avere la malattia. Questi dati supportano le conclusioni degli AA che in un soggetto con xerostomia e scialo-RM positiva la biopsia labiale per le ghiandole salivari minori può essere evitata.

Si noti che i valori delle probabilità post-test calcolate nei nostri esempi essenzialmente coincidono con i valori del VPP, che in fondo è il parametro del test legato alla prevalenza della malattia e che più immediatamente dà l'idea della probabilità di malattia nel singolo paziente.

Occorre però un'osservazione. Questi dati sono ricavati da valori di prevalenza piuttosto elevati ed è quello che si verifica nei Centri dove convergono determinate patologie su cui si applicano i test diagnostici specifici. Pertanto prima di dedurre le stesse probabilità di malattia sui propri pazienti è bene che ogni medico verifichi la prevalenza della malattia in studio nel suo Centro o area di lavoro ricavando dati di prevalenza dalla Letteratura Scientifica e da casistiche personali.

I concetti di LR e di probabilità post-test possono chiarire il fatto che un test diagnostico positivo per una malattia con bassa frequenza possa non dare una probabilità elevata di malattia nel singolo paziente. Infatti un basso valore di prevalenza comporterà un odds pre-test tendente allo 0, per cui il prodotto con cui si ricava l'odds post-test sarà tendenzialmente basso, a meno che il test non abbia una sensibilità e una specificità che si avvicinano al 100% per cui il LR sarà alto e questo favorirà una probabilità post-test elevata.

Il likelihood ratio e i test non dicotomici

La valutazione mediante sensibilità e specificità si può eseguire solo per test "dicotomici", cioè quando la risposta è del tipo presente/assente, normale/patologico. Quando il risultato di un test varia in una gamma di valori da normali a francamente patologici con valori intermedi che possono essere più o meno significativi per la diagnosi di patologia, occorre stabilire un valore cut-off al di sotto del quale il test è considerato negativo, viceversa positivo se il risultato è superiore. Un esempio di ciò è riscontrabile nell'articolo riportato nell'esempio 2 in cui gli esiti della scialoRM e della scintigrafia delle ghiandole salivari venivano classificati con uno score variabile da 0 a 4, i valori negativi corrispondevano ad uno score = 0, mentre quelli positivi avevano uno score ≥ 1 . Il LR risulta utile nell'analisi dei test non dicotomici, in quanto è possibile calcolare valori per diverse soglie di risultato del test.

Pertanto è possibile creare una scala di LR che vanno teoricamente da 0 all'infinito. Dal momento che la moltiplicazione del LR con gli odds pre-test dà gli odds post-test (e conseguentemente le probabilità post-test) è intuitivo che un LR pari a 1 non aggiunge alcuna informazione addizionale rispetto alla stima iniziale di probabilità di malattia; valori superiori a 1 aumentano la probabilità di malattia e, viceversa, valori inferiori a 1 la diminuiscono.

Nella tabella VI riportiamo range di valori del LR e il loro relativo impatto sulla probabilità di malattia.

LR+	LR-	Cambiamento della probabilità di malattia
> 10	< 0.1	ALTO
5-10	0.1 - 0.2	MODERATO
2-5	0.2 - 0.5	LIEVE
< 2	> 0.5	MINIMO o NULLO

Tab. VI: cambiamenti di probabilità di malattia in funzione del valore del LR

Tornando ai nostri esempi, nell'esempio 1 il LR+ è di poco superiore a 5, quindi ci aspetteremo un cambiamento moderato di probabilità di malattia nel singolo paziente valutato. Infatti la probabilità post-test è di 87%, ovvero 36,5% in più rispetto alla probabilità pre-test.

Nell'esempio 2 il LR+ per la sRM è molto alto (72,5), quindi il cambiamento di probabilità di malattia è alto, infatti la probabilità post-test è quasi del 100%; invece per la ScGS il LR+ è poco più di 1,7, quindi il cambiamento atteso è poco o nullo, in altre parole di fronte ad un soggetto con xerostomia tale esame potrebbe non essere dirimente nella diagnosi e non dare informazioni di rilievo.

Nella fig. 1 è riportato il nomogramma di Fagan che consente una rapida valutazione della probabilità post-test senza eseguire la serie di calcoli

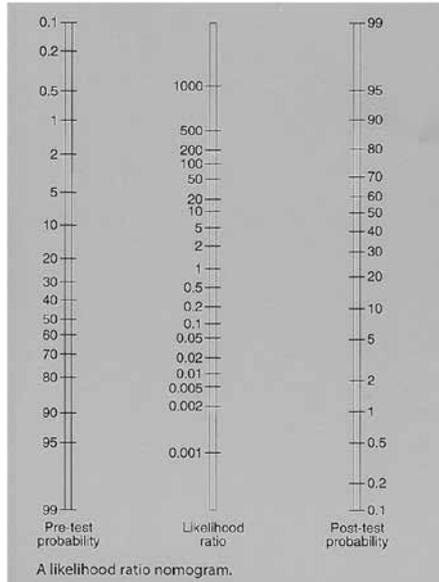


Fig. 1: nomogramma di Fagan per calcolare la probabilità post-test conoscendo la probabilità pre-test e il LR+. Con un righello unire i punti del pre-test e del LR+, l'incrocio del righello sulla terza linea dà il valore della probabilità post-test

sopra esposti. È sufficiente conoscere la probabilità pre-test e il LR+ (ricavabile come abbiamo detto conoscendo sensibilità e specificità). Con un righello si uniscono i due valori noti rispettivamente sulla prima e seconda linea. L'incrocio del righello sulla terza riga dà direttamente il valore della probabilità post-test.

Il likelihood ratio e le sequenze di test diagnostici

Il LR può essere utile nel calcolo della probabilità post-test quando si deve applicare una batteria di test diagnostici nello studio di una patologia. In altre parole la probabilità post-test del primo esame diventa la probabilità pre-test per il secondo test, che verrà moltiplicata per il LR del secondo test e così via. Proviamo ad esemplificare quanto detto.

Supponiamo di voler stabilire la diagnosi di OSAS grave in soggetti obesi (i dati che seguono sono di pura invenzione, a solo scopo esemplificativo). Per la nostra diagnosi scegliamo un valore antropometrico: l'indice di massa corporea (o BMI) e la saturimetria notturna.

Dalla nostra ipotetica indagine su 200 soggetti con BMI superiore a 25 emerge che i soggetti con un BMI superiore a 30 risultano effettivamente affetti da OSAS grave in 80/100 casi (= sensibilità), mentre soggetti con BMI inferiore a 30, ma superiore a 25 non risultano affetti da OSAS nel 75% dei casi (= specificità). Il LR+ per il BMI > 30 sarebbe di $0,80/(1-0,75) = 3,2$. A questo punto sapendo che la probabilità pre-test è del 50% (100 soggetti/200 con BMI superiore a 30) possiamo calcolarci la probabilità post-test:

$$\text{odds pre-test} = p/(1-p) = 0,50/0,50 = 1$$

$$\text{odds post-test} = \text{LR} \times \text{odds pre-test} = 3,2 \times 1 = 3,2$$

$$\text{probabilità post-test} = \text{odds post-test}/(1+\text{odds post-test}) = 3,2/4,2 = 0,76.$$

Ora consideriamo la saturimetria notturna. Ipotizziamo che una desaturazione sotto il 90% di almeno il 5% del tempo di registrazione abbia una sensibilità del 90% e una specificità dell'85%.

$$\text{Il LR+ per la saturimetria è di } 0,90/(1 - 0,85) = 6.$$

A questo punto ripetiamo i calcoli di sopra, utilizzando come probabilità pre-test il valore della probabilità post-test del BMI > 30:

$$\text{odds pre-test} = 0,76/(1 - 0,76) = 0,76/0,24 = 3,17$$

$$\text{odds post-test} = \text{LR} \times 3,17 = 6 \times 3,17 = 19,02$$

$$\text{probabilità post-test} = 19,02/(1 + 19,02) = 19,02/20,02 = 0,95.$$

Quindi applicando questa serie di 2 test, uno rilevato dall'esame obiettivo generale e l'altro da un esame strumentale di screening (saturimetria notturna), avremo la probabilità del 95% che un soggetto con BMI > 30 e con saturimetria notturna positiva per il 5% di desaturazione < 90% sia

affetto da OSAS grave. Ribadiamo che questi dati sono puramente immaginari, ma sperimentabili nella realtà clinica utilizzando diverse soglie dei test (diversi valori di BMI e diverse percentuali di desaturazione < 90%). In realtà esiste anche un sistema più rapido di calcolare la probabilità post-test finale: è sufficiente infatti moltiplicare l'odds pre-test iniziale con il LR_{tot} . Il LR_{tot} è dato dalla moltiplicazione di ogni LR dei vari test impiegati ($LR_{tot} = LR_1 \times LR_2 \times LR_3 \times \dots$).

Nel nostro esempio avremo:

odds pre-test = $0,50/0,50 = 1$. $LR_{tot} = 3,2 \times 6 = 19,2$.

Odds post-test = $1 \times 19,2 = 19,2$. Probabilità post-test finale = $19,2 / (1+19,2) = 19,2/20,2 = 0,95$.

Curva della caratteristica operativa di ricezione (COR)

Un altro modo per analizzare i test diagnostici non dicotomici è quello di utilizzare la curva della caratteristica operativa di ricezione o **ROC curve (Receiver Operator Characteristics)**, così definita perché derivata dall'addestramento delle radariste inglesi (vedi box n.° 2)

Una *ROC curve* esprime il tasso dei veri positivi (TVP), cioè la sensibilità, in funzione del tasso dei falsi positivi (TFP), che corrisponde al complemento della specificità (= $1 - TVN$).

Per capire meglio ciò che esprime una *ROC curve* osserviamo la fig. 2.

In alto sono rappresentate con curve gaussiane le distribuzioni dei due sottogruppi della popolazione studiata con un test diagnostico: malati (curva vuota), sani (curva piena). In basso sono raffigurate le relative *ROC curves*, tracciate su un sistema di assi cartesiani dove in ascisse sono riportati i valori del TFP (o $1 - \text{specificità}$) e in ordinate il TVP (o sensibilità), che di fatto sono gli stessi parametri utilizzati per il LR+.

In alto a sinistra è rappresentata una situazione ideale di test perfetto, un test cioè che consenta di tenere separati i due sottogruppi, individuando correttamente i soggetti ammalati. In questo caso vi è un cut-off ben definito al di sopra del quale tutti i soggetti sono veri positivi, al di sotto invece i soggetti sono veri negativi cioè soggetti sani.

Nella parte centrale invece è rappresentata la situazione di un test senza valore, in questo caso le due curve si sovrappongono; ciò sta ad indicare

Box n.° 2. Nota storica. Durante la 2ª Guerra Mondiale in Inghilterra le addette al radar venivano addestrate a riconoscere gli aerei nemici e distinguerli da quelli della propria nazione. Coloro che avevano la performance migliore, ovvero avevano la migliore ROC curve (alto numero di veri positivi e pochi falsi positivi), venivano selezionate per i radar posti alla difesa di Londra.

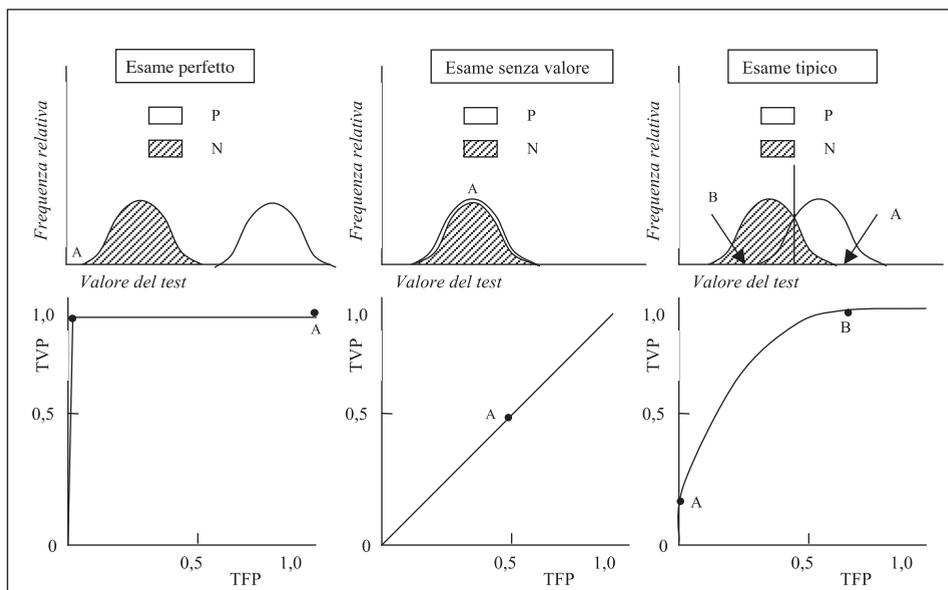


Fig. 2: distribuzione dei risultati d'esame e ROC curves di vari esami. I grafici superiori mostrano le funzioni della densità di probabilità per vari risultati d'esame di pazienti ammalati (P) e sani (N). I grafici inferiori mostrano le corrispondenti curve ROC.

che vi è il 50% di possibilità di individuare un soggetto malato o un soggetto sano. Di fatto questa situazione è paragonabile al lancio della moneta. La figura in alto a destra invece rappresenta la situazione tipica che realmente si riscontra: vi è una parziale sovrapposizione delle curve. In questa area di sovrapposizione un soggetto può appartenere al gruppo dei sani o dei malati, ma è erroneamente identificato dal test: avremo pertanto i falsi positivi o i falsi negativi. Va da sé che quanto minore è l'area di sovrapposizione tanto più ci avviciniamo ad una situazione ideale di test perfetto e viceversa.

Ora se immaginiamo di spostarci con una linea verticale perpendicolare all'asse delle ascisse (TFP) da destra verso sinistra, cambieremo progressivamente i valori soglia del nostro test (da valori molto alti a valori molto bassi). Se consideriamo le figure in alto, la nostra linea incontrerà prima i soggetti malati (VP), successivamente incrocerà l'area di sovrapposizione, ma adesso la nostra soglia (cut-off) del test è più bassa e quindi considererà soggetti malati anche alcuni soggetti sani (FP). Spostandosi ancora verso sinistra con l'approssimarsi dello zero verranno reclutati come positivi anche soggetti francamente sani.

In pratica una ROC curve si costruisce in questo modo: ogni volta che viene reclutato un vero positivo, la curva salirà verticalmente, sull'asse

delle ordinate o parallela ad essa, quando viene reclutato un falso positivo la curva si sposta orizzontalmente (aumenta il valore delle ascisse). Quindi tornando alla nostra figura 2, il grafico in basso a destra raffigura una tipica ROC curve: fino al punto A (valore soglia alto) sono reclutati solo veri positivi, da questo punto compariranno anche falsi positivi, pertanto la curva si sposta dalle ordinate, pur continuando a salire. Giunti al valore B la curva assumerà un plateau e proseguirà parallela alle ascisse in quanto tutti i veri positivi sono stati già reclutati, ma aumentano progressivamente solo i falsi positivi.

Nel grafico centrale, poiché i due gruppi si sovrappongono, la nostra curva salirà progressivamente per uguali valori di TVP e TFP e coinciderà con la retta bisettrice del piano cartesiano. Questa linea è chiamata anche *linea di equivalenza*.

Nel grafico a sinistra (situazione ideale) la prima parte della curva coincide con le ordinate fino al valore di 1,0, da questo punto decorre parallela alle ascisse (tutti falsi positivi) fino a raggiungere il corrispettivo valore di 1,0. In linea generale quanto più la curva è vicina alla linea di equivalenza tanto meno efficace è il test e viceversa. Questo concetto è espresso da un altro parametro che è *l'area sotto la curva* o *area under the curve* (AUC). Questo parametro assume valori da **0,5** a **1,0**. Un AUC di 0,5 significa che la curva coincide con la linea di equivalenza e quindi il test diagnostico non ha nessun valore pratico. Un valore prossimo a 1,0 indica un test molto efficace. Indicativamente il valore del test può essere classificato in base al valore dell'AUC nel seguente modo:

- 0,50 – 0,60 = inutile
- 0,60 – 0,70 = scarso
- 0,70 – 0,80 = discreto
- 0,80 – 0,90 = buono
- 0,90 – 1,00 = eccellente.

Una ROC curve risulta molto utile quando si devono valutare diverse soglie di un test diagnostico e scegliere il valore soglia che dà i migliori risultati.

Un esempio ci può chiarire meglio questo aspetto.

Supponiamo di valutare sensibilità e specificità di diverse soglie del parametro IT5 dei potenziali evocati acustici eseguito in un gruppo di soggetti affetti da neurinoma dell'acustico del diametro di circa 2 cm e in un gruppo di soggetti sani. Precisiamo che i dati che riportiamo sono di pura invenzione e servono solo a scopo esemplificativo.

Nella tabella VII riportiamo i dati relativi a 100 soggetti affetti da neurinoma dell'acustico (M+) e a 100 soggetti sani (M-).

Con i dati sopra riportati possiamo calcolare per ogni valore soglia il

Soglia IT5: 0,05 msec		M +	M -	TOT
	T +	95	90	185
	T -	5	10	15
	TOT	100	100	200
Soglia IT5: 0,10 msec		M +	M -	TOT
	T +	92	60	152
	T -	8	40	48
	TOT	100	100	200
Soglia IT5: 0,20 msec		M +	M -	TOT
	T +	85	40	125
	T -	15	60	75
	TOT	100	100	200
Soglia IT5: 0,30 msec		M +	M -	TOT
	T +	75	20	95
	T -	25	80	105
	TOT	100	100	200
Soglia IT5: 0,40 msec		M +	M -	TOT
	T +	40	5	45
	T -	60	95	155
	TOT	100	100	200

Tab. VII: tabelle 2 x 2 per il calcolo delle proprietà invarianti del test ABR per diverse soglie di IT5 (dati inventati a solo scopo esemplificativo)

TVP (sensibilità) e il TFP (1 – specificità); i valori sono riportati nella tabella VIII.

Ora riportiamo i dati su un sistema di assi cartesiani (fig. 3), dove in ascisse riportiamo i valori di TFP e in ordinata i valori di TVP. Il punto 1 corrisponde al valore soglia di IT5 = 0,05 msec, il punto 2 a IT5 = 0,10 msec, il punto 3 a IT5 = 0,2 msec, il punto 4 a IT5 = 0,3 msec e il punto 5 a IT5 = 0,4 msec. Congiungiamo poi i punti ottenuti partendo dallo zero con una linea curva e otteniamo la nostra *ROC curve*.

Per coloro che non sono in rapporto conflittuale con la matematica, ma anzi non disdegnano di assaporarla di tanto in tanto si rimanda la box

	Valori soglia di IT5				
	0,05 msec	0,10 msec	0,20 msec	0,30 msec	0,40 msec
TVP	0,95	0,92	0,85	0,75	0,4
TFP	0,90	0,60	0,30	0,20	0,05

Tab. VIII: valori di TVP e TFP per diversi valori soglia di IT5 per costruire la ROC curve

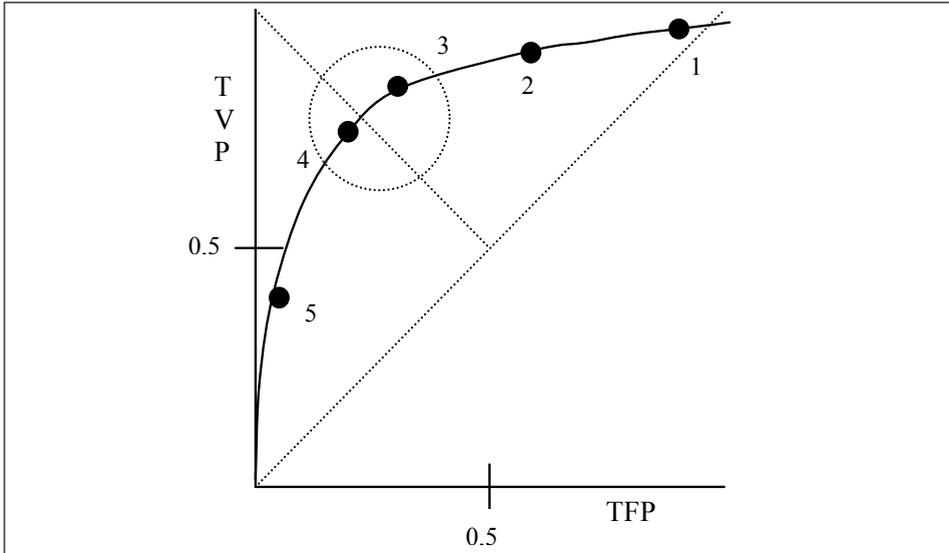
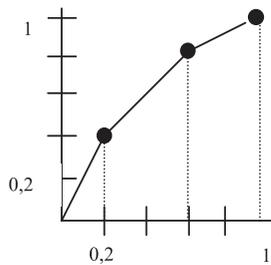


Fig. 3: ROC curve dell'ABR per il parametro IT5 su pazienti con neurinoma dell'acustico di circa 2 cm (esempio immaginato).

n.° 3 per un metodo semplice di calcolo della AUC. Qui ci limitiamo ad una valutazione grossolana osservando semplicemente la posizione della nostra curva rispetto alla linea di equivalenza, utilizzando una misura "empirica": il punto più distante della curva corrisponde a circa 3/5 della linea perpendicolare alla linea di equivalenza tracciata dall'angolo supe-

Box n°3. Calcolo della AUC



Osservando la figura a lato supponiamo che i tre punti abbiano coordinate 0,2 e 0,4; 0,6 e 0,8; 1 e 1. La linea spezzata (curva) e le normali condotte dai punti all'ascisse individuano un triangolo e 2 trapezi rettangolari. Pertanto è sufficiente calcolare l'area di queste figure e sommarle.

Area triangolo: $(0,2 \times 0,4)/2 = 0,04$
 Area 1° trapezio (al centro): $[(base\ maggiore + base\ minore) \times altezza]/2 = [(0,8 + 0,4) \times (0,6 - 0,2)]/2 = 0,24$
 Area 2° trapezio (a dx): $[(1 + 0,8) \times (1 - 0,6)]/2 = 0,36$
 Sommando le aree otteniamo la nostra AUC:
 $0,04 + 0,24 + 0,36 = \underline{0,64}$

riore sinistro. Assumendo che $1/5$ di questa linea corrisponda a circa $0,1$ di AUC, l'AUC è approssimativamente di $0,8$ e quindi potremo giudicare il nostro test discreto – buono.

Quando si ricerca una soglia ottimale con una *ROC curve* questa solitamente corrisponde al tratto della curva più distante dalla linea di equivalenza (tratto segnato nel cerchio nel grafico di fig. 3). In questi punti si ottengono il miglior TVP con il più basso TFP.

Proviamo ora ad applicare le *ROC curves* ai nostri esempi di prima.

ESEMPIO N.° 1.

Dai dati esposti in precedenza sappiamo che la sensibilità è di $0,869$ (= $86,9\%$) e il TFP (= $1 - \text{specificità}$) è di $0,133$. Tracciamo il punto nel nostro grafico e uniamo con una curva lo zero, il nostro punto e l'angolo in alto a destra ($1,0; 1,0$) (fig. 4). La curva risulta lontana dalla linea di equivalenza e ciò indica che il test è buono. Va precisato però che in questo caso abbiamo utilizzato un test dicotomico: esame citologico benigno/maligno e pertanto abbiamo solo una coppia di valori TVP/TFP. Il decorso della curva è solo presunto.

ESEMPIO N.° 2.

Più interessante è l'applicazione delle *ROC curves* all'esempio n.° 2 (fig. 5). Gli AA ci forniscono i valori dei VP per ogni stadio radiologico della SS dei due esami studiati. Nella tabella IX abbiamo sintetizzato i valori di sensibilità per ogni stadio.

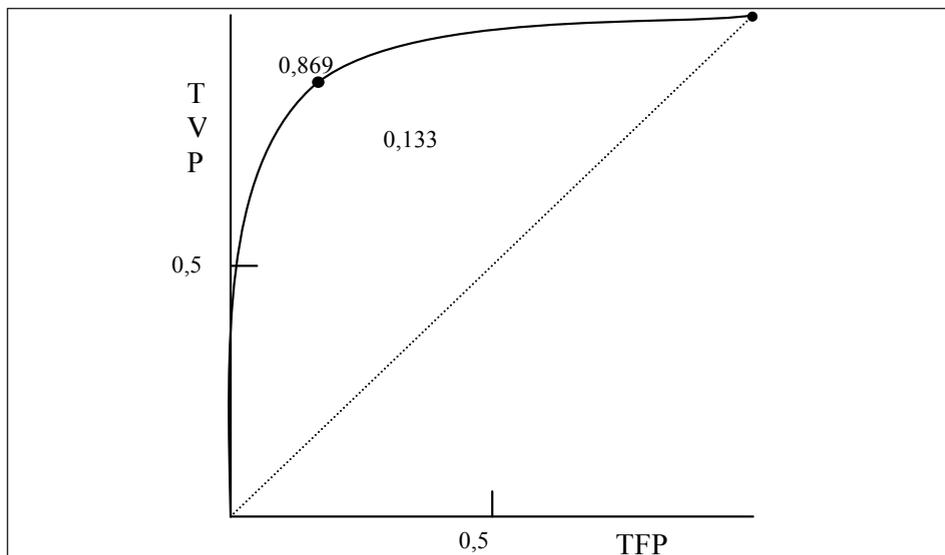


Fig. 4: ROC curve per l'esempio n.° 1

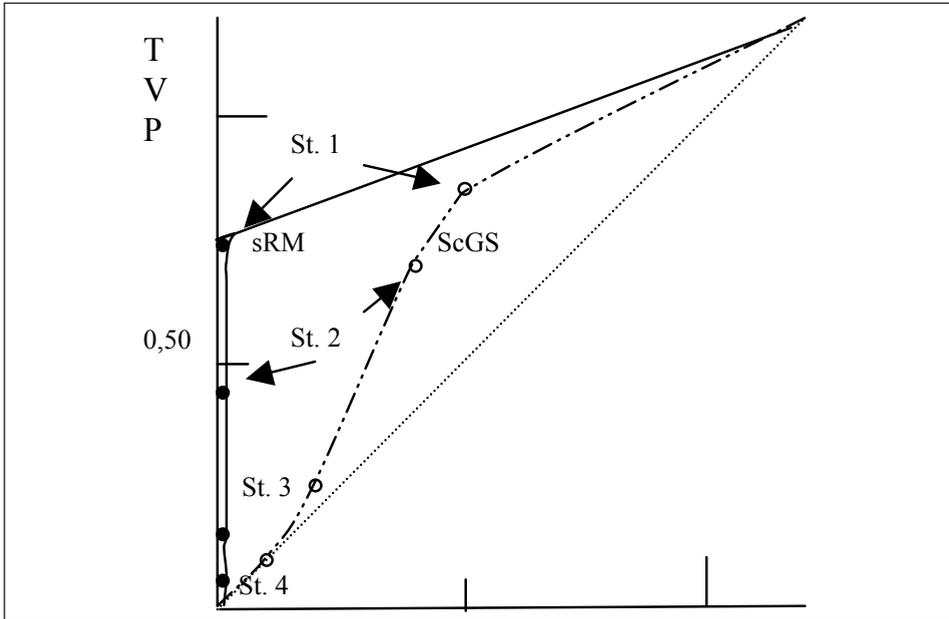


Fig. 5: ROC curves applicate all'esempio 2

Per la sRM il valore del TFP per i vari stadi rimane 0 (sensibilità 100%) in quanto tale valore è già presente alla soglia più bassa (stadio 1) e pertanto aumentando la soglia del test i falsi positivi possono solo diminuire e non aumentare. Per la ScGS la specificità per gli stadi superiori a 1 non è desumibile dall'articolo, pertanto i valori sono solo puramente indicativi a scopo esemplificativo supponendo appunto che il TFP tende a ridursi con l'aumentare della soglia del test.

Riportando i valori nel solito sistema di assi cartesiani (fig. 5) otteniamo le 2 ROC curves, la linea tratteggiata indica quella per la ScGS, la continua

Scialo RM				
	Stadio 1	Stadio 2	Stadio 3	Stadio 4
TVP	0,73	0,43	0,15	0,05
TFP	0	0	0	0
Scintigrafia delle Gh. Salivari				
	Stadio 1	Stadio 2	Stadio 3	Stadio 4
TVP	0,86	0,71	0,26	0,1
TFP	0,5	0,4*	0,2*	0,1*

Tab. IX: valori di TVP e TFP per i diversi stadi radiologici della SS dedotti dall'articolo dell'es. 2. (*) valori supposti a scopo esemplificativo

	RA +	RA -	Tot.
SFAR \geq 7	108	20	128
SFAR < 7	38	103	141
Tot.	146	123	269

Tab. X: tab. 2x2, dati relativi allo studio citato nel testo. RA: rinite allergica. SFAR: score for allergic rhinitis

quella per la sRM. Se osserviamo la distanza delle curve dalla linea di equivalenza, deduciamo che la ScGS è un test tendenzialmente scarso nella diagnosi di SS.

La *ROC curve* per la sRM non raggiunge i valori di eccellenza, ma si può esprimere comunque un giudizio di test buono per la SS, anche se, come visto in precedenza, la sRM raggiunge valori di probabilità post-test elevatissimi.

Per concludere questa trattazione riportiamo un altro esempio di applicazione clinica pratica dell'utilizzo della *ROC curve* per stabilire un cut off diagnostico. L'esempio è desunto dal seguente articolo: Annesi-Maesano I, Didier A, Klossek M et al. *The score for allergic rhinitis (SFAR): a simple and valid assessment method in population studies*. *Allergy* 2002; 57: 107-114. In questo lavoro gli Autori (AA) hanno elaborato un questionario per la diagnosi della rinite allergica (RA). Il questionario è costituito da alcuni items che riguardano sintomatologia, familiarità, precedente diagnosi di RA e precedenti prove allergiche, allergeni scatenanti. Lo score (SFAR) va da 0 ad un massimo di 16. Sono stati studiati 269 pazienti che afferivano alle strutture di riferimento per problemi nasali. Questi pazienti sono stati sottoposti a visita specialistica e alcuni di essi anche al prick test. Gli AA hanno eseguito una validazione diagnostica calcolando sensibilità, specificità, VPP, VPN, indice di Youden e costruendo una *ROC curve* per ogni valore dello score. Hanno trovato che il *cut off* significativo corrisponde al valore di SFAR \geq 7. Per questo valore gli AA riportano i seguenti indici di performance del test:

- sensibilità 74%
- specificità 83%
- valore predittivo positivo 84%
- valore predittivo negativo 74%.

Il questionario validato è stato somministrato telefonicamente ad un campione di 3001 soggetti di cui il 21% è risultato avere uno score > 7. Rielaboriamo ora le analisi. Costruiamo una tabella 2 x 2 tale da ottenere i valori sopra elencati (tab. X). Come già si può osservare la probabilità pre-test è molto alta pari al 54%.

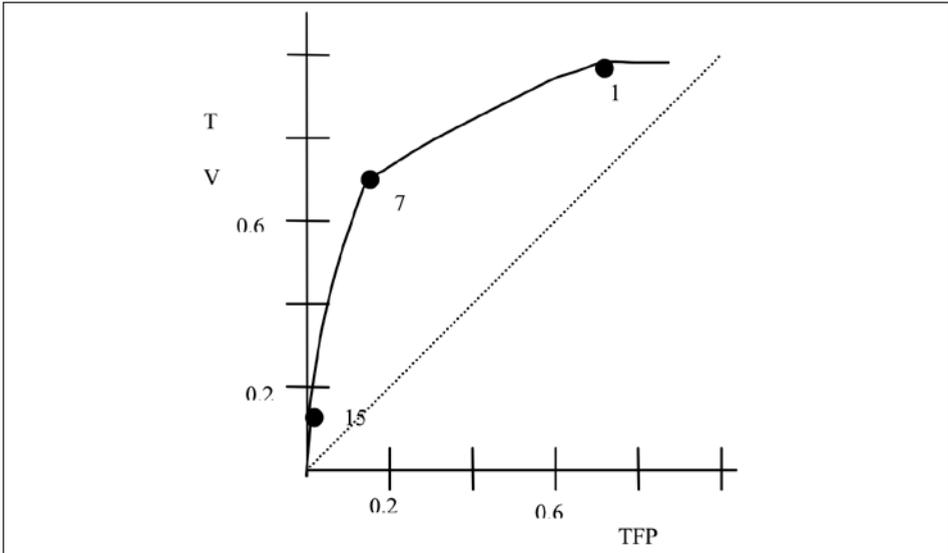


Fig. 6: ROC curve relativi ai dati dello studio citato nel testo. I numeri della curva indicano il punteggio dello score (SFAR).

Calcoliamo ora il LR+ (sensibilità/(1 - specificità) = 0,74/(1 - 0,83) = 0,74/0,17 = 4,35; valore abbastanza discreto.

Con questi parametri possiamo risalire alla probabilità post-test o utilizzando il nomogramma di Fagan o calcolandola come indicato nei paragrafi precedenti, il valore che si ottiene è 84%, corrispondente al VPP, un valore che possiamo ritenere senz'altro buono, ma partendo da una probabilità pre-test alta.

Per questa soglia del test diagnostico (SFAR = 7) l'indice di Youden (J) corrisponde a:

$$J = \text{sensibilità} + \text{specificità} - 1 = 0,74 + 0,83 - 1 = 0,57.$$

Gli AA hanno trovato che questo è il valore più alto rispetto a quelli ottenibili con gli altri punteggi dello SFAR.

Gli AA hanno costruito una ROC curve ottenuta per ogni valore di TVP (= sensibilità) e TFP (= 1 - specificità) di ogni singolo punteggio dello score. Riportiamo in forma modificata la curva nella fig. 6. Come si osserva più è basso il valore dello SFAR più è alto il valore del TFP, ovvero è alto il numero di falsi positivi (in pratica si includono sia soggetti allergici che non allergici). Viceversa più è alto lo score più basso è il TFP, ma risulta basso anche il TVP. Ovvero con un punteggio massimale o quasi massimale dello score si reclutano soggetti sicuramente allergici, ma risulta troppo restrittivo perché verrebbero inclusi solo i soggetti che rispondono positivamente a quasi tutti gli items. Quindi avremo elevata speci-

ficità, ma bassa sensibilità perché il cut off troppo alto farebbe passare una buona quota di soggetti effettivamente malati nei falsi negativi (per comprendere meglio questo concetto rivedere la tab. VII relativi ai dati simulati dell'ABR). A questo punto non resta che calcolare l'AUC (per un calcolo approssimativo si può utilizzare il sistema indicato nel box n.° 3). Il valore che si ottiene è poco inferiore a 0,8. Il test si può definire discreto. Lasciamo al lettore le considerazioni sulla applicabilità di questo score, come test diagnostico, sulla propria popolazione di riferimento.

Bibliografia

- Zoccali C. *Le basi statistiche ed epidemiologiche della medicina clinica*. [FULL TEXT]
www.accmed.net/stat/libro/index.htm
oppure da www.gimbe.org > Home page > Bibliografia EBM > Pubblicazioni in lingua italiana > (testo citato).
- Tape TG. *Interpreting Diagnostic Test*.
<http://gim.unmc.edu/dxtests/Default.htm>
Simon S. *Stats: ROC curve*.
www.cmh.edu/stats/ask/roc.asp
- Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-based Medicine. How to Practice and Teach EBM*. ELSEVIER CHURCHILL LIVINGSTONE, 2005, 3th Edition.
- Bliss D, Diggs B, Matar M. *Improving Diagnostic Accuracy: Simple Statistical Nomograms to Interpret Medical Literature*. World J Surg 2010 Apr 20 [Epub ahead of print].
- Coulthard MC. *Quantifying how tests reduce diagnostic uncertainty*. Arch Dis Child 2007; 92: 404-8.

ANALISI CRITICA DELLE PROCEDURE TERAPEUTICHE

S. Palma. V. Mele

Una volta nel Medioevo si credeva che i Re di Francia e Inghilterra avessero poteri taumaturgici e potessero guarire le malattie, in particolare le scrofole (linfadeniti tubercolari ulcerate e suppuranti), con il tocco delle mani ¹.

Immaginiamo che alcuni medici del Medioevo avessero voluto constatare la veridicità di questo fatto applicando una metodologia scientifica ispirata ai principi dell'EBM (medicina basata sulle prove di efficacia). Essi avrebbero dovuto procedere nel seguente modo.

Avrebbero dovuto selezionare i soggetti malati (esempio soggetti affetti da scrofole del collo), suddividerli a caso (random) in due gruppi, ad esempio assegnando ad essi un numero ogni volta diverso e sorteggiare a caso i numeri, assegnando alternativamente i soggetti estratti nei due gruppi; avrebbero poi dovuto verificare l'omogeneità dei gruppi per quanto riguarda le caratteristiche generali (percentuale di soggetti di sesso maschile, età media, durata della malattia). Successivamente avrebbero dovuto scegliere due personaggi "taumaturghi", uno dei quali effettivamente sarebbe stato il Re, l'altro un simulatore. Un gruppo di malati sarebbe stato presentato ad un personaggio e il secondo gruppo all'altro personaggio, con l'accortezza però che né i malati né i medici sapessero chi effettivamente fosse il Re. Successivamente tutti i malati sarebbero stati seguiti per un certo periodo di tempo sufficiente per verificare la guarigione (probabilmente alcuni mesi). Alla fine dello studio i risultati sarebbero stati semplicemente riassumibili come numero di soggetti guariti nei due gruppi, rivelando quelli toccati dal Re (gruppo trattato) e quelli toccati dal simulatore (gruppo controllo). Se effettivamente il Re avesse avuto poteri taumaturgici ci si sarebbe aspettato un numero di guariti nettamente superiore nel gruppo trattato rispetto al gruppo controllo.

Per noi medici moderni non è difficile immaginare quale sarebbe stato il reale risultato!

Questo esempio, che non ha la pretesa di processare la Storia, illustra quali sono i passaggi necessari per ottenere le prove di efficacia dei trattamenti terapeutici. Il disegno di studio necessario per ottenere ciò è tipico

¹ Marc Bloch, *"I re taumaturghi"*, Einaudi, Torino 1989.

degli studi controllati e randomizzati (RCTs: Randomized Controlled Trials). Per valutare la reale efficacia di un trattamento ed evitare errori metodologici e di interpretazione dei risultati (*biases*) è necessario procedere come segue:

- allocare a caso i pazienti arruolati in due gruppi che devono risultare omogenei per caratteristiche specifiche legate alla malattia; infatti se i due gruppi differiscono per stadio della malattia e il farmaco sperimentato agisse con maggiore efficacia nelle fasi iniziali della malattia e venisse somministrato a soggetti che si trovano nello stadio iniziale della malattia, mentre i soggetti del gruppo controllo si trovano in una fase più avanzata, è ovvio che i risultati, prevedibilmente positivi, non saranno reali come quelli che si osserveranno quando il farmaco verrà utilizzato nella popolazione generale. Il gruppo di controllo si rende necessario anche per ovviare ai possibili risultati positivi legati alla evoluzione favorevole della malattia trattata: se il farmaco è realmente efficace il numero di guarigioni deve essere più alto nel gruppo trattato rispetto al controllo.
- I due gruppi devono risultare anche omogenei per caratteristiche generali, in particolare per età e sesso. Infatti ci possono essere farmaci che possono dare risposte diverse nei due sessi, o in funzione dell'età (esempio nel soggetto anziano la risposta può essere più torbida oppure possono manifestarsi più facilmente degli effetti collaterali).
- La scelta dei pazienti da arruolare nello studio deve essere eseguita mediante criteri ben definiti e chiari. Per essere ammessi allo studio i soggetti devono avere determinate caratteristiche e la malattia diagnosticata con criteri chiari e scientificamente validi (criteri di inclusione), mentre non devono possedere altre caratteristiche che non possono andare bene per lo studio in questione, come ad es. l'età minore, la gravidanza, ecc. (criteri di esclusione). Una volta che i soggetti hanno rispettato i criteri di inclusione e di esclusione, vengono arruolati e inseriti a caso (random) in uno dei due bracci dello studio.
- Procedere in *doppio cieco*. Ad un gruppo di soggetti viene somministrato il trattamento in studio (solitamente un farmaco), all'altro un *placebo* (sostanza inerte) o un trattamento di controllo già validato; ma né i pazienti né i medici o gli sperimentatori sanno quale dei due trattamenti stanno seguendo. Questo è importante soprattutto nella valutazione dei risultati da parte degli sperimentatori, perché se sono a conoscenza del trattamento, c'è la tendenza inconscia e anche in buona fede ad enfatizzare i risultati positivi del trattamento in questione, o renderli positivi se sono borderline, viceversa a sot-

tostimare i risultati nel gruppo controllo (bias di misurazione e di interpretazione).

- Gli obiettivi che si vogliono perseguire nello studio devono essere ben definiti (endpoints). Esempio se lo studio è condotto su un antibiotico l'endpoint principale sarà la guarigione di una infezione (tonsillite, otite ecc.). Per quanto riguarda il concetto di endpoint surrogato vedi cap. 3. Si possono stabilire e valutare altri endpoints secondari e complementari, nell'esempio in questione potrebbero essere il numero di recidive di infezioni in un determinato arco di tempo o il numero di effetti collaterali maggiori e minori. Sugli endpoints così definiti devono essere valutati i risultati utilizzando parametri di valutazione appropriati. Nel caso di un processo infettivo si riterrà raggiunto l'obiettivo (paziente guarito) quando ad esempio non sarà più presente febbre, sarà scomparso il dolore, non ci saranno più i segni clinici di infezione acuta (iperemia, edema, essudato). Già da questo esempio emerge il problema di come nella pratica clinica non sia sempre facile stabilire un confine tra normale e patologico, soprattutto quando la valutazione è soggettiva e legata al giudizio di un osservatore, che può differire da quello di un altro. Pertanto gli sperimentatori per ovviare in parte a questo problema stabiliscono una gradazione (score) dei vari sintomi e segni che si vogliono valutare, indicando solitamente con 0 nessun cambiamento o la situazione peggiore e con il numero più alto, variabile a seconda della scala utilizzata, la situazione migliore possibile o il massimo cambiamento ottenibile. Viene stabilito poi un cut-off al di sotto del quale il risultato è ritenuto negativo (es. non guarito o non migliorato), viceversa al di sopra di esso positivo (es. guarito o migliorato). Questo modo di definire i risultati è importante per valutarne la significatività clinica, mentre per valutare la significatività statistica ci si avvale dei test più appropriati, da scegliere a seconda delle misure utilizzate e del tipo di studio. La valutazione statistica spesso non dà idea di quanti pazienti effettivamente beneficiano del trattamento, ma solo una idea generale dell'efficacia del trattamento.
- Altro punto da rispettare e da definire bene è il *follow-up*. Infatti i risultati attesi del trattamento sperimentato devono essere valutati dopo un adeguato intervallo di tempo. Ad esempio potremo valutare l'effetto di un analgesico nel post operatorio già dopo poche ore dalla sua somministrazione, valutando l'analgesia a intervalli variabili es. 6, 12, 24 ore. Ovviamente non avrà senso valutarne l'effetto di un giorno di trattamento dopo una settimana. Al contrario se vogliamo valutare l'efficacia di un antibiotico, raccoglieremo

i dati preferibilmente dopo pochi giorni o una settimana e non nelle prime ore dopo la somministrazione di una o due dosi. A seconda del tipo di trattamento sperimentato il follow-up può durare settimane o anche mesi; nel caso di chemioterapici anche anni.

Da quanto finora esposto risulta chiaro che i RCTs sono un ottimo strumento per valutare l'efficacia dei farmaci. Meno bene si prestano per valutare l'efficacia dei trattamenti chirurgici. Innanzitutto alcuni interventi chirurgici, come quelle di tipo oncologico o di carattere urgente, non possono essere sottoposti a una valutazione di questo tipo, randomizzata e con un gruppo controllo, per ovvi motivi deontologici. La loro efficacia piuttosto si basa sull'esperienza clinica condivisa e trasmessa tra i vari medici, e sui risultati chirurgici osservati in senso longitudinale (tipo studi di coorte). Da questo tipo di analisi emergono più facilmente i fattori prognostici e di rischio.

Gli interventi chirurgici che possono essere valutati con studi controllati e randomizzati sono quelli di elezione, la cui mancata esecuzione non inficia sostanzialmente il benessere del paziente. In ambito otorinolaringoiatrico l'esempio classico è quello condotto sugli interventi di adenotonsillectomia in pazienti con sintomatologia moderata (non problemi ostruttivi, non recidive frequenti di faringo-tonsilliti) che potrebbero comunque beneficiare di terapia medica (a tal proposito si consiglia la lettura dell'articolo di Van Staaïj BK et al.: *Effectiveness of adenotonsillectomy in children with mild symptoms of throat infections or adenotonsillar hypertrophy: open, randomised controlled trial*. BMJ, 10.09.2004). In questo caso un gruppo verrebbe indirizzato alla chirurgia, l'altro a terapia medica o semplicemente osservato nel tempo. In ogni caso la valutazione della terapia chirurgica non può essere condotta *in cecità*: ovviamente il chirurgo deve sapere chi opera e che intervento fare e il paziente inevitabilmente è a conoscenza del trattamento a cui viene sottoposto e nel caso di intervento chirurgico deve dare il suo consenso (si parla allora di *studio randomizzato aperto*). In questi casi quindi è più opportuno che vi sia un osservatore esterno che valuti i risultati, per evitare il condizionamento da parte del chirurgo, che tenderebbe ad enfatizzare i propri risultati.

Durante la conduzione di un RCT ci sono alcune accortezze da rispettare. Innanzitutto è importante evitare "*la contaminazione*" del trattamento con interventi esterni. Può accadere infatti che un paziente arruolato assuma un altro farmaco di sua iniziativa o su prescrizione di altro medico che può interferire nel processo di guarigione della malattia, bisogna allora registrare tutte le interferenze esterne che devono essere prese in considerazione nella valutazione finale. L'interferenza può essere tale che il

paziente debba uscire dal trial (drop out). L'uscita può avvenire anche per decisione del paziente sia durante il trattamento che alla fine non presentandosi ai controlli del follow-up.

Importante poi è valutare la *compliance* dei pazienti arruolati al trattamento imposto. Infatti se il trattamento risultasse sgradevole (eccessivi effetti collaterali, somministrazione parenterale dolorosa, ecc.) è probabile che si verifichi un eccessivo abbandono di soggetti dal trial, soprattutto dal gruppo in cui viene sperimentato il farmaco, con ovvie distorsioni sull'analisi dei risultati, anche per la possibile perdita della "cecità", in quanto i soggetti del gruppo trattato verrebbero facilmente identificati. Un aspetto importante dell'analisi dei risultati di un RCT è che questa deve essere condotta sui pazienti con "*intenzione a trattare*" (*intention to treat: ITT*). Per questo aspetto si rimanda al cap. 4.

L'analisi dei risultati

Come possono essere valutati i risultati?

Per analizzare i risultati di un RTC torna utile la tabella 2 x 2, analogamente a quella impiegata per i test diagnostici (vedi anche cap. 1).

Bisogna però tenere presente che i risultati di un trattamento derivati da uno studio RCT vengono intesi come riduzione del **rischio** di malattia. Esempio, se abbiamo a che fare con un farmaco antivertiginoso, ci si aspetta che ci sia una buona percentuale di risoluzione della vertigine, ma i risultati sono analizzati in termine di rischio di malattia, cioè di persistenza della sindrome vertiginosa.

Pertanto nella prima colonna della nostra tabella 2 x 2 (vedi Tab. I), si riporta il numero di soggetti che presentano ancora la malattia (non guariti, indicati con M+), nella prima riga il numero relativo al gruppo trattato, nella seconda quello relativo al gruppo controllo. Nella seconda colonna invece riporteremo il numero di soggetti che risultano guariti o perché sottoposti al trattamento o per risoluzione spontanea della malattia. Questi saranno indicati con M-.

Con il termine di **rischio assoluto (RA)** si intende la probabilità che un individuo presenti l'evento studiato (es. malattia) nel corso di un periodo

	M +	M-	Totali
Trattati	a	b	a + b
Controlli	c	d	c + d
Totali	a + c	b + d	Popolazione studiata

Tab. I: tabella 2 x 2 per l'analisi dei risultati dei RCTs

di tempo specificato. Con riferimento alla nostra tabella corrisponde al rapporto tra soggetti M+ e il totale del gruppo in esame (tasso di eventi nel gruppo trattato o di controllo):

$$RA = \frac{\text{n. eventi}}{\text{n. pazienti}}$$

ovvero:

Tasso di eventi nel gruppo trattato = Experimental event rate (EER) = $a/(a + b)$;

Tasso di eventi nel gruppo di controllo = Control event rate (CER) = $c/(c + d)$.

Il valore di questo rapporto è compreso tra 0 e 1.

Con il termine di **rischio relativo (RR)** si intende il rapporto tra i due rischi assoluti ovvero il rischio di un gruppo esposto ad un evento (es. trattamento farmacologico) relativamente ad un altro di confronto non esposto all'evento in studio (gruppo controllo):

$$RR = \frac{\frac{\text{n. eventi (gr. Trattati)}}{\text{n. pazienti (gr. Trattati)}}}{\frac{\text{n. eventi (gr. Controllo)}}{\text{n. pazienti (gr. Controllo)}}$$

Pertanto esprimeremo il rischio relativo nel seguente modo:

Rischio relativo = RelativeRisk (RR) = $[a/(a + b)]/[c/(c + d)]$.

La **riduzione del rischio relativo (RRR)** misura la riduzione percentuale del rischio del verificarsi di un evento nel gruppo dei non esposti o dei trattati rispetto agli esposti o ai non trattati. Esso si calcola nel seguente modo:

Riduzione del rischio relativo = Relative Risk Reduction (RRR) = $(CER - EER)/CER = [c/(c + d)] - [a/(a + b)]/ [c/(c + d)]$.

A questo punto chiariamo questi concetti finora espressi con un semplice esempio immaginario (dati di pura invenzione). Supponiamo di analizzare un grosso RCT in cui 1000 bambini con OME sono stati seguiti per almeno un anno valutando il rischio di otiti ricorrenti (gruppo controllo), mentre altrettanti bambini sono stati sottoposti a profilassi antibiotica per tre mesi e seguiti per lo stesso intervallo di follow-up (gruppo trattato). Il risultato positivo considerato è di ≤ 2 episodi di OMA/anno. Se nel gruppo controllo il numero di bambini con otiti ricorrenti è stato di 200 (CER = 200/1000) e nel gruppo trattato di 120 (EER = 120/1000), il rischio relativo è di $0,120/0,200 = 0,6$ (60%). In altre parole se attribuiamo al gruppo controllo un rischio standard di

1,0, il gruppo trattato (bambini sottoposti a profilassi antibiotica) ha un rischio relativo di 0,6 cioè più basso. La riduzione del rischio relativo ci dice in che percentuale si è ridotto il rischio relativo di otiti ricorrenti nel gruppo trattato. Dalla formula prima esposta possiamo calcolare il nostro RRR:

$$(0,200 - 0,120)/0,200 = 0,80/0,200 = 0,40 \text{ (40\%)}$$

Quindi nel gruppo trattato la riduzione del rischio relativo per le otiti ricorrenti è del 40%. Lo stesso risultato lo otteniamo se semplicemente sottraiamo al rischio standard (1,0) il rischio relativo (0,6).

Per il clinico risulta più facilmente comprensibile ed immediatamente utilizzabile un altro parametro, che è la **riduzione del rischio assoluto (RRA)**, espresso semplicemente come differenza tra i due rischi assoluti (del gruppo trattato e del gruppo controllo).

La sua espressione sarà:

$$\text{Riduzione del rischio assoluto} = \text{Absolute Risk Reduction (ARR)} = \text{CER} - \text{EER} = [c/(c + d)] - [a/(a + b)].$$

Utilizzando ancora l'esempio di prima, la profilassi antibiotica in 1000 bambini con OME riduce il numero di otiti ricorrenti da 200 a 120; la riduzione di rischio assoluto per le otiti ricorrenti sarà dato dalla seguente sottrazione: $200 - 120 = 80$ eventi ($= 200/1000 - 120/1000 = 0,2 - 0,12 = 0,08 = 8\%$).

Parlando di efficacia di un trattamento il termine di rischio può generare confusione. Se ad esempio stiamo valutando l'efficacia di un antibiotico nella risoluzione di otiti medie acute, dovremo intendere che il "rischio" di persistenza dell'infezione verosimilmente sia più alto nel gruppo controllo. In altre parole facendo riferimento alla tabella 2 x 2 se l'antibiotico in questione è veramente efficace ci aspetteremo valori di "b" e di "c" alti e quindi il valore del CER tenderà a 1, mentre il valore di EER tenderà a 0. Dal valore di ARR si ricava un parametro molto importante che è il **numero di pazienti da trattare**, che si esprime semplicemente come l'inverso dell'ARR:

$$\text{Numero di pazienti da trattare} = \text{Number Needed to Treat (NNT)} = 1/\text{ARR} =$$

$$= \frac{1}{[c/(c + d)] - [a/(a + b)]}$$

Questo parametro indica quanti pazienti si devono sottoporre al trattamento studiato per avere 1 paziente beneficiato o prevenire 1 evento sfavorevole. Nella condizione ideale con il CER vicino a 1 e l'EER vicino a 0, il rapporto sopra indicato tenderà a 1; ciò indicherebbe una situazione ideale in quanto avremmo a disposizione un trattamento quasi perfetto: praticamente una guarigione per ogni paziente trattato.

Va da sé che un NNT alto è indice di un trattamento poco efficace o francamente scarso.

Attualmente il valore di NNT è il parametro più potente a disposizione per misurare la “significatività clinica” (= efficacia) di un trattamento.

Volendo calcolare il NNT nell'esempio di prima avremo: $1/0,08 = 12,5$. Cioè bisogna trattare circa 12 – 13 bambini per prevenire un caso di otiti ricorrenti.

Ricordiamo che $1/0,08 = 10/0,8 = 100/8 = 1000/80$. In altre parole il NNT può essere calcolato dividendo il numero di soggetti a rischio (nel nostro esempio 1000) per il numero di successi (80).

Manteniamo ancora questo esempio fittizio, ma ipotizziamo che nel gruppo dei bambini trattati il numero di casi con otiti ricorrenti non sia 120, ma 300, più alto del gruppo controllo.

In questo caso l'ARR sarà: $0,2 - 0,3 = -0,1$ e quindi il NNT avrà un valore negativo pari a -10 . Quando il NNT assume un valore negativo vuol dire che il trattamento ha un effetto dannoso, in questi casi si parla **Number Needed to Harm (NNH)**, cioè numero di casi necessario per osservare un evento avverso. Il suo reciproco non è più l'ARR, ma è definito **Absolute Risk Increase (ARI)**. Il NNH può essere utilizzato per valutare quantitativamente un danno (effetti collaterali) prodotto da un trattamento. Contrariamente al NNT è auspicabile che il NNH sia alto, ovvero siano necessari molti pazienti trattati per osservare un evento avverso.

Una misura che viene impiegata negli studi caso-controllo e soprattutto nelle metanalisi è l'**odds ratio (OR)**, che altro non è che una stima del rischio relativo, ossia il rischio di avere un certo evento (malattia) tra i casi trattati o esposti ad una noxa e il rischio di avere lo stesso evento tra i controlli (non trattati o non esposti). Nel caso delle metanalisi, per ciascun studio esaminato, l'OR misura la probabilità che un risultato si verifichi nel gruppo trattato diviso la probabilità che si verifichi nel gruppo non trattato.

Prima di calcolare l'OR si ricordi che l'*odds* (vedi anche cap. 5) è un rapporto tra due probabilità espresso matematicamente come $p/(1 - p)$.

Riferendoci all'esempio di sopra se nei bambini non trattati il rischio di otiti ricorrenti è del 20% (200/1000), in termini di odds il rischio è espresso come $200/800 = 1:4 = 0,25$.

L'OR è quindi un rapporto tra due odds. Facendo riferimento alla tabella 2 x 2 (Tab. I) per il gruppo dei trattati il “*p*” corrisponde ad “*a*” e “ $1 - p$ ” corrisponde a “*b*”, mentre per il gruppo controllo il “*p*” corrisponde a “*c*” e “ $1 - p$ ” corrisponde a “*d*”. A questo punto possiamo calcolare l'OR come segue:

$$\text{Odds Ratio (OR)} = (a/b)/(c/d) = (a/b) \times (d/c) = \underline{(a \times d)/(b \times c)}$$

Se il valore di OR è inferiore a 1, vuol dire che la probabilità dell'evento è inferiore nel gruppo di trattamento, ovvero il risultato è favorevole al trattamento. Un risultato pari a 1 indica che le probabilità sono pari in entrambi i gruppi. Un valore superiore a 1 indica che la probabilità è maggiore nel gruppo di trattamento e quindi si tratta di un risultato sfavorevole al trattamento.

Se volessimo calcolare l'OR per l'esempio formulato prima, avremo:

$$OR = (120 \times 800) / (200 \times 880) = 96000 / 176000 = 0,54,$$

dove 800 e 880 sono i soggetti M- rispettivamente del gruppo controllo e del gruppo trattato.

ESEMPIO N.° 1

Per chiarire i concetti finora espressi analizziamo un paio di articoli presi dalla Letteratura Scientifica in modo da fornire degli esempi pratici. Il primo è un interessante articolo che analizza l'efficacia del Ginkgo biloba nel trattamento degli acufeni (Rejali D, Sivakumar A, Balani N. *Ginkgo Biloba does not benefit patients with tinnitus: a randomized placebo-controlled double-blind trial and meta-analysis of randomized trials*. Clin Otolaryngol 2004; 29: 226-231).

In questo articolo gli AA riportano i risultati di un proprio studio RCT condotto su pazienti sofferenti di acufeni per valutare l'efficacia del Ginkgo biloba e i risultati di una metanalisi condotta su studi analoghi. La metanalisi comprende oltre allo studio degli AA altri 5 studi che presentavano caratteristiche e metodologie analoghe. La metodologia e gli obiettivi sono espressi molto chiaramente. Gli AA hanno valutato il disturbo dei pazienti somministrando ad essi prima e dopo il trattamento due questionari: il Tinnitus Handicap Inventory (THI) e il Glasgow Health Status Inventory (GHSI), come valutazione complementare hanno utilizzato l'audiometria tonale valutando la media delle frequenze 0,5, 1, 2 e 4 kHz per entrambi gli orecchi. Il THI costituiva la misurazione principale (endpoint principale).

Lo studio è stato condotto in doppio cieco, un gruppo di pazienti ha assunto il Ginkgo biloba per os (120 mg una volta die) e l'altro ha assunto il placebo. Il trattamento è durato 12 settimane e la valutazione dei risultati è stata eseguita alla fine del trattamento. Non tutti i pazienti arruolati (33 per gruppo) hanno terminato il trattamento. L'analisi è stata condotta sui pazienti che hanno terminato il trattamento (31 del gruppo trattato, 29 del gruppo controllo).

Data la numerosità piuttosto limitata di questo studio gli AA hanno eseguito una metanalisi per ottenere un campionario più ampio e ottenere una maggiore significatività dei dati complessivi.

Per brevità riportiamo nella nostra tabella 2 x 2 la somma dei dati dei sei lavori selezionati dagli AA. (Tab.II).

	M +	M-	Totali
Trattati	415	107	522
Controlli	417	87	504
Totali	832	194	1026

Tab.II: tabella 2 x 2 che riporta i dati della metanalisi dello studio citato nell'esempio (vedi testo)

Come si osserva dalla tabella i soggetti complessivamente studiati sono stati 1026, di cui 522 appartenevano al gruppo trattato e 504 al gruppo controllo. Il numero di soggetti che ha beneficiato dal trattamento non è molto cospicuo, circa un quinto per ogni gruppo. Procediamo ora con la nostra analisi:

$$\text{CER} = 417/504 = 0,83$$

$$\text{EER} = 415/522 = 0,80$$

$$\text{RR} = 0,80/0,83 = 0,96$$

$$\text{RRR} = (0,83 - 0,80)/0,83 = 0,04$$

$$\text{ARR} = 0,83 - 0,80 = 0,0323$$

$$\text{NNT} = 1/0,0323 = 30,9$$

$$\text{OR} = (415 \times 87)/(107 \times 417) = 0,81.$$

Da questi dati emerge subito che il numero di eventi (soggetti che non sono migliorati) nei due gruppi è percentualmente alto (circa 80%), ma il dato più interessante che riassume le conclusioni dello studio è il NNT molto elevato (circa 31). In pratica questa metanalisi ci dice che per ottenere un paziente beneficiato dal trattamento con Ginkgo biloba bisogna trattare circa 31 pazienti!

Ora se questi dati possono essere applicati ai nostri pazienti per affinità etniche, territoriali e nosologiche possiamo dedurre che non vi è alcuna utilità nell'impiegare nella terapia degli acufeni il Ginkgo biloba.

Come nota dobbiamo segnalare che gli AA riportano un valore di OR uguale a 1,236; questo valore si ottiene se nella formula sopra riportata si invertono il dividendo e il divisore.

Nell'articolo in esame troviamo un altro parametro che non abbiamo accennato: l'**intervallo di confidenza (Confidence Interval CI)**.

L'intervallo di confidenza, inteso solitamente al 95% (95%CI), è un intervallo di valori al cui interno si trova con una probabilità del 95% il valore medio del parametro preso in esame. In altre parole ripetendo l'esperimento 100 volte, dovremo osservare per almeno 95 volte il valore del parametro in esame compreso nel range definito dall'intervallo di confidenza.

Il valore di CI è inversamente proporzionale alle dimensioni del campione

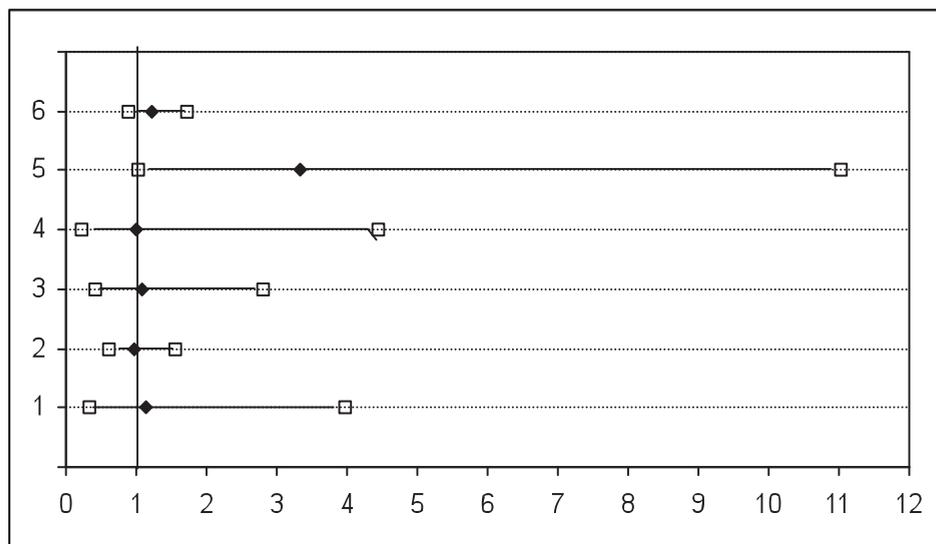


Fig. 1: diagramma di Peto, sviluppato dai dati riportati nell'articolo citato (vedi testo). Non riportato un lavoro in quanto fuori scala (Morgenstern & Biermann 2002). In ordinata riportati i seguenti lavori citati: 1) AA dell'articolo citato; 2) Drew & Davies 2001; 3) Meyers 1986; 4) Holgers et al 1994; 5) Morgenstern & Biermann 1997; 6) Sintesi della metanalisi.

utilizzato nello studio. Quando il campione è costituito da un numero di soggetti limitato, il range del CI risulta piuttosto ampio.

Dal punto di vista clinico gli intervalli di confidenza forniscono una misura quantitativa dell'entità clinica dell'effetto e pertanto sono più informativi del test di significatività statistica, che dà invece un risultato di soglia (statisticamente significativo sì/no).

Nelle metanalisi solitamente l'OR e il CI dei vari lavori vengono rappresentati graficamente con il così detto *diagramma di Peto*. Nella figura 1 abbiamo costruito un diagramma di Peto con i dati esposti nella tabella 6 del citato articolo. Abbiamo mantenuto i valori indicati e per semplicità di rappresentazione abbiamo escluso un lavoro che presentava valori di infinito. La linea verticale centrata sull'uno rappresenta una linea di demarcazione per i valori dell'OR, indicati nel grafico con piccoli rombi pieni: se il valore si colloca a sinistra allora i risultati sono a favore del trattamento, se invece si collocano a destra sono sfavorevoli al trattamento. Il CI è rappresentato da una linea orizzontale, i cui estremi nel nostro grafico sono rappresentati da piccoli quadrati vuoti, nel cui centro si trova il valore dell'OR. Quanto più lunga è la linea tanto più è ampio il CI, ciò indica un campionamento ridotto e una scarsa significatività dei risultati. Se poi la linea del CI incrocia la linea verticale dell'uno, i risultati dello

studio indicano una scarsa efficacia del trattamento. Come si osserva nella figura tutti gli studi indicano una scarsa efficacia della terapia degli acufeni con Ginkgo biloba. La sintesi della metanalisi (n° 6 in ordinata) mostra un CI breve, che è quello auspicabile con le metanalisi in quanto la numerosità dei soggetti studiati è alta, ma si colloca a cavallo della linea dell'uno con un OR prossimo a 1, rafforzando così le conclusioni dei singoli studi.

ESEMPIO N.° 2

Il secondo articolo prende in esame l'efficacia del *Cineolo* (o *Eucaliptolo*), principale ingrediente dell'olio di Eucalipto che avrebbe effetti antinfiammatori e trova indicazioni nelle comuni patologie "da raffreddamento" (Kehrl W, Sonnemann U, Dethlefsen U. *Therapy for Acute Nonpurulent Rhinosinusitis With Cineole: Results of a Double-Blind, Randomized, Placebo-Controlled Trial*. *Laryngoscope* 2004; 114: 738-42). Gli AA hanno selezionato 150 pazienti affetti da rinosinusite non purulenta. La diagnosi era basata sulla valutazione clinica otorinolaringoiatrica e sulla sintomatologia, ad ogni sintomo veniva dato un punteggio in base alla gravità (0 assenza, 1 lieve, 2 moderato, 3 grave). Il totale del punteggio dava lo score sintomatologico. Eleggibili erano i pazienti con uno score superiore a 10; il valore massimo dello score era 25. I pazienti sono stati suddivisi in maniera randomizzata in 2 gruppi, ciascuno di 75 pazienti: il primo gruppo (trattati) riceveva il farmaco (2 capsule di gelatina con 100 mg di Cineolo 3 volte al dì prima dei pasti), il secondo gruppo (controllo) riceveva il placebo (un preparato con le stesse caratteristiche organolettiche) con la stessa modalità di somministrazione. A tutti i pazienti di entrambi i gruppi veniva prescritto come decongestionante nasale la xilometazolina spray 3 volte al dì. Il trattamento è durato 7 giorni, la valutazione è stata fatta al 4° e 7° giorno (follow-up). L'endpoint primario era la valutazione del miglioramento clinico misurato come riduzione dello score sintomatologico globale. Gli endpoint secondari erano la valutazione della variazione dello score dei singoli sintomi e la tollerabilità della terapia (valutazione degli effetti collaterali).

Ci soffermiamo sul risultato principale ovvero quanti pazienti hanno effettivamente beneficiato della terapia. Gli AA ci riportano come misura del miglioramento clinico la riduzione di oltre il 50% dello score sintomatologico rispetto al valore iniziale pretrattamento (baseline).

Questo è un dato molto importante, infatti in molti articoli che riportano l'efficacia di una terapia si ritrova solo il dato statistico (solitamente significativo) basato sul confronto tra valore globale post-trattamento e valore pre-trattamento (*valutazione intragruppo*) e sul confronto del valore post-trattamento con il gruppo controllo, talvolta neanche presente (*valuta-*

	M +	M-	Totali
Trattati	6	69	75
Controlli	55	20	75
Totali	61	89	150

Tab.III: tabella 2 x 2 che riporta i dati relativi allo studio del Cineolo in soggetti con rinosinusite non purulenta (vedi testo). Soggetti migliorati al 7° giorno di follow-up

zione intergruppo). Tuttavia in questi articoli non è dato sapere quanti pazienti siano effettivamente guariti o migliorati e se il dato è riportato manca il criterio obiettivo con cui viene definito il miglioramento clinico. Quindi si raccomanda di specificare sempre quale sia la misura obiettiva per definire un paziente “migliorato” o “guarito”. Come già detto in precedenza è importante quantificare il quadro clinico mediante uno score dei segni e sintomi in modo da uniformare i pazienti oggetto dello studio. Tornando al nostro articolo gli AA ci informano che al 7° giorno i pazienti non migliorati nel gruppo dei trattati erano 6, mentre nel gruppo controllo erano 55. A questo punto possiamo inserire i dati nella nostra tabella 2 x 2 ed analizzarli (tab. III).

I risultati che si ottengono sono i seguenti:

$$\text{CER} = 55/75 = 0,73$$

$$\text{EER} = 6/75 = 0,08$$

$$\text{RR} = 0,08/0,73 = 0,11$$

$$\text{RRR} = (0,73 - 0,08)/0,73 = 0,89$$

$$\text{ARR} = 0,73 - 0,08 = 0,65$$

$$\text{NNT} = 1/0,65 = \mathbf{1,53}$$

$$\text{OR} = (6 \times 20)/(69 \times 55) = 0,03.$$

I dati elaborati forniscono un NNT ottimale, ovvero sarebbe sufficiente trattare 3 pazienti per ottenere due beneficiati dalla terapia. Lasciamo al lettore la rivalutazione della metodologia utilizzata nello studio e la eventuale applicabilità della terapia con Cineolo ai propri pazienti.

Intervallo di confidenza per il NNT

Anche per il NNT è possibile calcolare l'intervallo di confidenza. Gli estremi dell'intervallo corrispondono ai reciproci dei rispettivi ARR. Nel box 1 è riportato il metodo di calcolo. Anche in questo caso l'ampiezza dell'intervallo di confidenza del NNT dà l'idea della reale validità del dato ottenuto. Quanto più stretto è l'intervallo con gli estremi vicini al valore di NNT ottenuto, tanto più significativo è il risultato e quindi tanto mag-

Box n.° 1. Calcolo del 95%CI del NNT.

Ponendo $p_1 = \text{EER}$, $p_2 = \text{CER}$, $n_1 = a + b$ e $n_2 = c + d$ (vedi Tab. I), si calcola per prima cosa l'errore standard SE:

$$SE = \sqrt{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2}$$

L'intervallo di confidenza al 95% per l'ARR si calcola nel seguente modo:

$$95\%CI \text{ ARR} = \text{ARR} \pm (1,96 \times SE)$$

da qui si ricava il 95%CI di NNT:

Limite inferiore 95%CI NNT:

$$1/[\text{ARR} + (1,96 \times SE)]$$

Limite superiore 95%CI NNT:

$$1/[\text{ARR} - (1,96 \times SE)]$$

giore sarà la validità clinica dello studio. Viceversa se l'intervallo è ampio l'impatto clinico dello studio sarà minore e il valore del NNT deve essere interpretato con giudizio considerando il tipo di terapia studiata, la patologia da trattare, la propria popolazione di pazienti di riferimento; in altre parole ogni medico deve valutare rischi e benefici attraverso il filtro della propria esperienza.

Applichiamo ora l'intervallo di confidenza del NNT ai risultati dei 2 esempi precedentemente esposti.

Nel primo lavoro otteniamo i seguenti valori: NNT 30,9; 95%CI: -64,7; 12,5.

Nel secondo lavoro i dati sono i seguenti: NNT 1,53; 95%CI: 1,3; 1,9.

I dati del secondo lavoro sono di facile interpretazione, gli estremi dell'intervallo sono entrambi positivi e sono vicini al valore centrale, l'intervallo è infatti abbastanza ristretto e questo è indice di bontà dello studio e della terapia.

I dati del primo lavoro sono di non semplice interpretazione, innanzitutto il valore inferiore è negativo e il valore del NNT non sembra essere compreso nell'intervallo come ci si aspetterebbe.

Come detto in precedenza un valore negativo di NNT indica un effetto "dannoso" ed è esprimibile come NNH eliminando il segno negativo. Intuitivamente possiamo comprendere che se il NNT è alto vuol dire che il trattamento ha scarsa efficacia e che la differenza tra gruppo trattato e gruppo controllo è minima, ed è possibile che somministrando una terapia sostanzialmente inefficace a più persone si possano anche osservare più effetti collaterali che benefici. D'altra parte con un trattamento inefficace ripetendo l'esperimento più volte è possibile rilevare in alcuni casi un numero maggiore di "beneficiati" nel gruppo controllo pur in assenza di effetti collaterali. In questi casi per semplice espressione matematica avremo un NNT negativo ovvero registreremo un NNH. Quando la differenza tra eventi nei due gruppi è minima la riduzione del rischio assoluto (ARR) tenderà a 0 e quindi il valore di NNT tenderà all'infinito (inteso qui come concetto matematico, non filosofico o astronomico!). Quando

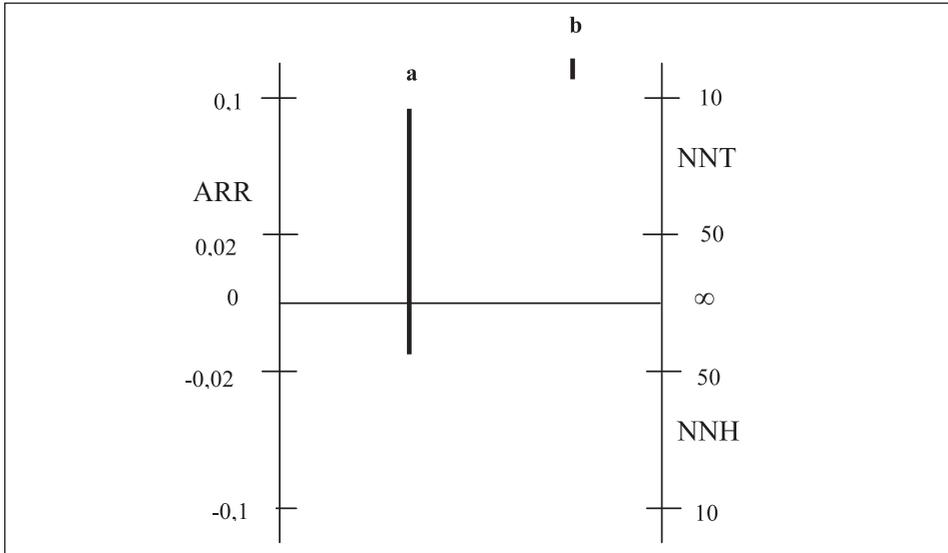


Fig. 2: Relazione tra ARR (asse delle ordinate a sn) e NNT (asse delle ordinate a dx). Per valori negativi il NNT diventa NNH. La transizione tra NNT e NNH avviene “attraverso l’infinito” (per $ARR = 0$). La linea “b” indica il 95%CI relativo allo studio del 2° esempio (sul Cineolo nelle rinosinuiti non purulente) e la linea “a” indica il 95%CI relativo allo studio del 1° esempio (sul Ginkgo biloba negli acufeni). Questo intervallo di confidenza “attraversa l’infinito”, ma comprende il valore di NNT 30,9.

invece l’ARR è massimale (100%) il valore di NNT sarà uguale a 1 (nel caso di $ARR = -100\%$ avremo $NNH = 1$). Quindi per valori molto piccoli di ARR il valore di NNT tenderà all’infinito e progressivamente “sconfinerà” nel NNH. Questo concetto può essere compreso meglio osservando la fig. 2. Nella figura la linea verticale indicata con “b” è molto breve ed è compresa tutta nel range del NNT tra 1 e 2, esprimendo l’ampiezza del 95%CI trovato per lo studio del 2° esempio. La linea “a” invece esprime il 95%CI relativo al primo studio, quello sul Ginkgo biloba nel trattamento degli acufeni, che risulta molto ampio tale da “attraversare l’infinito” e sconfinare nel NNH, ma come si può osservare sul grafico comprende il valore di 30,9 corrispondente al NNT trovato nello studio.

In parole povere uno studio ben fatto e corretto non solo riporta il valore di NNT, ma anche il suo intervallo di confidenza al 95%. Il lettore deve quindi considerare semplicemente tre aspetti:

1. valore del NNT: quanto più è piccolo, prossimo a 1, tanto più efficace risulta la terapia.
2. Ampiezza del 95%CI: quanto più è piccola con gli estremi vicini al valore di NNT tanto più significativo è il risultato dello studio. Solitamente uno studio condotto su un numero ampio di sog-

getti (adeguato campione), in presenza di una terapia efficace dà come risultati un NTT piccolo e 95%CI stretto; questo implica una reale evidenza scientifica e una applicabilità della terapia ai propri pazienti.

3. Valore degli estremi del 95% CI: se gli estremi sono molto distanti dal NNT e soprattutto se quello inferiore ha valore negativo è presumibile che la terapia sia inefficace se non addirittura dannosa. Non vi è evidenza scientifica che la terapia possa portare un reale beneficio ai pazienti trattati e i risultati positivi riscontrati nel gruppo controllo possono essere conseguenti all'evoluzione favorevole della malattia, ovvero guarigione o miglioramento spontaneo.

Bibliografia

- Zoccali C. *Le basi statistiche ed epidemiologiche della medicina clinica*. [FULL TEXT]
www.accmed.net/stat/libro/index.htm
oppure da www.gimbe.org > Home page > Bibliografia EBM > Pubblicazioni in lingua italiana > (testo citato).
- Cordell WH. *Number Needed to Treat (NNT)*. *Ann Emerg Med* 1999; 33(4): 433-436.
- Levin M, Walter S, Lee H, Haines T, Holbrook A, Moyer V, for the Evidence-Based Medicine Working Group. *How to Use an Article about Harm*.
www.cche.net/usersguides/harm.asp.
- Straus SE, Richardson WS, Glasziou P, Haynes RB. *Evidence-based Medicine. How to Practice and Teach EBM*. ELSEVIER CHURCHILL LIVINGSTONE, 2005, 3th Edition.
- Altman DG. *Confidence intervals for the number needed to treat*. *BMJ* 1998; 317: 1309-12.

FATTORI DI RISCHIO, COME INTERPRETARLI E COME ANALIZZARLI

C. Vicini, P. Canzi

L'epidemiologia contemporanea studia notoriamente la distribuzione delle malattie nella popolazione, dando risposta a tre quesiti cardine sintetizzati nell'espressione anglosassone "Who has what?" "When?" e "Where?" (Chi sono le persone malate? Dove e Quando la malattia si manifesta più frequentemente?). Storicamente, il ragionamento causale (che risponde alla domanda "Why?" – "Perché?") si è sviluppato progressivamente a partire dalle prime ipotesi di diffusione infettivologica di malattia (pioniere del modello causale in medicina è stato Robert Koch agli inizi del 1900). Il concetto di "*fattore di rischio o esposizione*" emerge solo successivamente con la scoperta delle malattie cronico-degenerative, quindi col *decadere dell'idea deterministica* causa-effetto tipica dei processi morbosi su base infettiva. La nuova relazione esaminata implica un ragionamento di tipo *probabilistico* in cui si riconosce che la presenza di un fattore aumenta la probabilità di comparsa della malattia, ma non necessariamente la implica. Analogamente, l'assenza del fattore di rischio non esclude la presenza della malattia, distinguendosi profondamente dal concetto deterministico di noxa patogena. Per esempio, il fumo di sigaretta è sicuramente un fattore di rischio per il carcinoma laringeo (in quanto ne aumenta la probabilità di comparsa), ma non tutti i malati di carcinoma laringeo hanno necessariamente fumato in passato. Si definiscono quindi fattori di rischio quei *fattori che sono associati positivamente con il rischio di sviluppare una malattia ma che non sono sufficienti a determinarla*. Più genericamente, i fattori di rischio si riferiscono ad *ogni agente in grado di influenzare lo stato di salute della persona*. Sebbene in letteratura sia facile il riscontro di svariate classificazioni dei fattori di rischio è peraltro possibile ricavarne tre categorie:

1. Fattori di rischio Individuali
2. Fattori di rischio Ambientali
3. Fattori di rischio Comportamentali

Ciascuna categoria abbraccia molteplici aspetti, di cui alcuni possono essere soggetti a modifica (es. le abitudini di vita) mentre altri sono da considerarsi imm modificabili (es. i fattori genetici).

Tra gli scopi principali dell'attività medico-scientifica vi è la ricerca dei fattori in grado di modificare lo sviluppo di una determinata malattia, nonché la stima quantitativa del loro potenziale effetto.

Nell'analisi di uno studio scientifico risulta essenziale avere gli strumenti per interpretarne i risultati prima di giungere alla conclusione di un reale rapporto causale fra fattore di rischio e malattia. Inevitabilmente anche i migliori trials clinici randomizzati possono incorrere in errori che, se non riconosciuti, inficiano negativamente le conclusioni del lettore.

Quattro quesiti dovrebbero sempre accompagnare lo studio di una relazione fra fattore di rischio e malattia:

- I. Esiste una relazione di *Causalità* fra Esposizione e Malattia?
- II. L'Associazione osservata fra Esposizione e Malattia può ricondursi ad "*Errori Sistemati*"?
- III. L'Associazione osservata fra Esposizione e Malattia può ricondursi a "*Fattori di Confondimento*"?
- IV. Esiste una relazione di *Causalità* fra Esposizione e Malattia?

Alcuni di questi aspetti sono già stati descritti nei capitoli precedenti, per cui ci limiteremo ad analizzare sinteticamente ciascuna domanda nell'ottica di una rapida ed agevole interpretazione del fattore di rischio.

I. Esiste una relazione di *Causalità* fra Esposizione e Malattia?

Per effetto del caso, l'esposizione e la malattia accadono insieme più frequentemente di quanto ci si possa attendere. Escludere una tale eventualità significa:

- a) Verificare l'esistenza di una "*associazione statistica*" fra esposizione e malattia
- b) Verificare l'esistenza di una "*associazione statistica significativa*" fra esposizione e malattia
- c) Calcolo dell'*intervallo di confidenza*
 - a) L' *associazione statistica* fra esposizione e malattia ne *esprime il grado di dipendenza*. La *Forza o Intensità* dell'associazione statistica *rappresenta la misurazione quantitativa* del grado di dipendenza. Non esiste un'unica procedura di calcolo per ottenere la Forza dell'associazione statistica. Di seguito ne riportiamo la formulazione più frequente e generica:

<div style="border: 1px solid black; padding: 5px; display: inline-block;"> Forza dell'Associazione Statistica </div>	=	$\frac{\text{Frequenza della Malattia nei soggetti Esposti}}{\text{Frequenza della Malattia nei soggetti Non Esposti}}$
---	---	---

Per chiarire il significato pratico di tale rapporto, si prendano in esame le seguenti condizioni limite:

- 1) *La frequenza di malattia nei soggetti esposti è uguale alla frequenza di malattia nei soggetti non esposti (Forza = 1).* È intuitivo che questa circostanza esprima la completa assenza di un'associazione statistica fra fattore di rischio e malattia.
- 2) *La malattia è presente solo nei soggetti esposti all'ipotetico fattore di rischio (Forza $\rightarrow \infty$).* In questo caso il fattore in esame potrebbe essere causa della malattia.
- 3) *La malattia è presente solo nei soggetti non esposti. (Forza = 0).* In questo caso il fattore potrebbe essere agente protettivo della malattia.

Chiaramente le condizioni esaminate sono del tutto ipotetiche, poiché nella realtà la malattia si manifesta in modo differente in entrambi i gruppi studiati (esposti e non esposti). Di conseguenza, due sono i possibili valori assunti dall'Intensità dell'associazione statistica:

- *Forza > 1 se la frequenza di malattia è maggiore negli esposti rispetto ai non esposti*
- *Forza < 1 se la frequenza di malattia è minore negli esposti rispetto ai non esposti*

Condizione necessaria perché sussista un'associazione statistica fra malattia ed esposizione è il riscontro di un grado di dipendenza con Forza > 1 .

- b) Nella progettazione di un lavoro scientifico è necessaria l'individuazione della popolazione di studio, che per motivi ovviamente pratici è rappresentata da un campione di individui e non dalla totalità dei soggetti che presentano la malattia in esame.

La *variabilità campionaria* è la variazione casuale che esiste fra campione e campione quando dalla stessa popolazione si estraggono a caso campioni diversi. Logicamente è proprio l'estrazione casuale di un campione ad imporre sin dal nascere un errore *non eliminabile, dettato dal caso*. Questo fenomeno è *riducibile solo incrementando la numerosità del campione studiato*.

L'*associazione statistica* calcolata fra esposizione e malattia (punto a) potrebbe essere quindi frutto del caso, da cui la necessità di elaborare specifici test di inferenza statistica che verifichino se i risultati debbano o non debbano attribuirsi alla variabilità campionaria.

I test di significatività consentono il calcolo di un valore numerico P, che esprime la possibilità di relazione casuale, o non, tra malattia ed esposizione. Convenzionalmente un valore di $P < 0.05$ esclude la possibilità che il caso possa

ritenersi spiegazione della relazione riscontrata. Il valore di P dipende sia dalle dimensioni del campione esaminato, sia dalla magnitudine dell'associazione statistica (vedi anche cap. 4).

Concludendo, si riconosce un'associazione statistica significativa solamente quando i test di significatività abbiano escluso la possibilità ($P < 0.05$) che le differenze osservate fra frequenza di malattia nei soggetti esposti versus non esposti ($F > 1$) sia attribuibile al caso.

c) Il concetto di *intervallo di confidenza* è ampiamente spiegato nel capitolo dell'Analisi delle procedure terapeutiche (cap. 6). Si ricorda, che al fine di poter estendere i risultati del campione esaminato all'intera popolazione, si rendono necessarie particolari stime statistiche intervallari utili a stabilire la significatività dell'effetto di un trattamento. L'*intervallo di confidenza* permette di essere nel giusto (con una certa probabilità pari al livello di confidenza, per esempio 95%) affermando che il vero valore del parametro studiato è compreso all'interno dell'intervallo stesso. Cerchiamo di chiarire un concetto di non facile intuizione. Per esempio, si ricava da uno studio su campione che il carcinoma dell'esofago è 6 volte maggiore nei forti bevitori rispetto ai bevitori occasionali di alcool con una $P < 0.05$ (associazione statistica significativa). Se si vogliono estendere i risultati all'intera popolazione è indispensabile il calcolo dell'intervallo di confidenza che poniamo essere compreso tra 3 e 18 ("IC = [3;18]"). Il valore ottenuto significa che, con un livello di confidenza del 95%, nella popolazione generale il rischio di carcinoma esofageo nei forti bevitori rispetto a quelli occasionali è verosimilmente compreso fra 3 e 18. Quindi il valore 6 rappresenta unicamente la misura più verosimile in termini assoluti, ma nella popolazione generale il rischio può andare da un minimo di 3 fino ad un massimo di ben 18 volte.

L'*intervallo di confidenza* fornisce quindi maggiori informazioni rispetto alla semplice P e dipende principalmente dalla numerosità del campione studiato: quanto maggiore è la numerosità del campione, tanto più stretto è l'intervallo di confidenza e quindi la precisione della stima di confidenza (il concetto di intervallo di confidenza è dunque legato alla potenza dello studio – vedi anche cap. 4).

È d'obbligo ricordare che i metodi di valutazione statistica verificano unicamente il fenomeno della variabilità campionaria e non possono in alcun

modo controllare l'errore dovuto ai *Bias* o *Errori Sistemati* ed ai *Fattori di Confondimento*.

II. L'Associazione osservata fra Esposizione e Malattia può ricondursi ad "Errori Sistemati"?

Il termine anglosassone "*Bias*" significa letteralmente "Distorsione". In ambito statistico la parola resta in traducibile ed è sinonimo di "*Errore Sistemato*" riferendosi agli *errori che alterano in modo costante i risultati di uno studio scientifico*. Gli "Errori Sistemati o Bias" sono dunque *eliminabili con una corretta programmazione della ricerca* poiché esito di un errore noto che può essere presente in qualsiasi fase della progettazione o conduzione del lavoro.

Sebbene ne esistano di differenti tipi, nella pratica tre sono le categorie di Bias da conoscere:

1. Bias di Selezione
2. Bias di Misurazione o di Informazione
3. Performance Bias

1. I *Bias di Selezione* sono errori sistemati conseguenti ad un *erroneo campionamento dei soggetti appartenenti allo studio*, implicando differenze fra le caratteristiche degli inclusi allo studio rispetto ai non facenti parte. I Bias di selezione producono quindi un campione di studio non rappresentativo della popolazione a cui si intendono estendere i risultati della ricerca. Costituiscono un problema rilevante negli studi caso-controllo e negli studi trasversali, sebbene possano influenzare anche i risultati degli studi di coorte e dei trials clinici randomizzati.

Analizziamo i più frequenti Bias di Selezione:

- Campionamento non corretto. Si attribuisce ad una sbagliata definizione della popolazione a cui si intendono estendere i risultati dello studio, oppure ad una non corretta estrazione del campione. Ad esempio, immaginiamo di voler realizzare una ricerca sui fattori di rischio nei pazienti con epistassi avendo come popolazione campione i degenti in un determinato ospedale. Ovviamente i pazienti con epistassi che ricorrono all'ospedalizzazione sono scarsamente rappresentativi di tutte le persone affette da epistassi, perché ne esprimono solo la componente più grave. Qualora i risultati dello studio venissero estesi a tutti coloro che hanno riportato nella loro vita almeno un episodio di epistassi, si commetterebbe un Bias di Selezione da campionamento non corretto.

- Inclusione di volontari nel campione. L'inclusione di soggetti volontari nel campione di studio è soggetta al rischio di selezionare persone con specifiche motivazioni per il loro stato di salute.
- Bias di intervista. Sono riconducibili:
 - a) *al soggetto intervistato:* la ricerca di informazioni riguardanti specifici fattori di rischio potrebbe essere influenzata da particolari motivazioni del paziente. Se ad esempio si desidera indagare l'esposizione alla cocaina, il paziente potrebbe mentire nel riferirne l'assunzione.
 - b) *all'intervistatore:* se l'intervistatore conosce l'appartenenza dell'intervistato al gruppo dei casi o dei controlli, potrebbe inconsciamente influenzarne le risposte.

2. I *Bias di Misurazione o di Informazione* sono errori sistematici conseguenti a classificazioni o metodi non corretti, di misurazione dell'esposizione o della malattia. L'introduzione di tali errori può essere ricondotta all'osservatore, all'oggetto di studio oppure allo strumento adoperato per le misurazioni.

A seconda della modalità con cui si realizza l'interferenza fra fattore di rischio (o malattia) e misurazione, si riconoscono due tipologie di Misclassificazioni (errori di classificazione dell'esposizione o della malattia):

Misclassificazione Non Differenziale: è dovuto all'utilizzo di misure non corrette che incidono in modo analogo sia sul gruppo dei casi che su quello dei controlli. In altre parole l'errore di classificazione (Misclassificazione) del fattore di rischio (o della malattia) non è associato alla malattia (o al fattore di rischio) ed è dunque detto "non differenziale". In questa circostanza, tutti i soggetti appartenenti allo studio hanno la stessa probabilità di subire un Bias di Misurazione e la Forza dell'associazione statistica tende ad 1 (Forza \rightarrow 1 o "diluizione della misura di effetto dell'esposizione").

Immaginiamo ad esempio di voler studiare il rapporto fra alcool e carcinoma del cavo orale. A scopo semplificato distinguiamo l'esposizione al fattore di rischio (alcool) in due sole categorie: abuso di alcool, assunzione occasionale. Per ovvi motivi, alcuni dei soggetti che abusano cronicamente di alcool possono riferirne un'assunzione occasionale. Come conseguenza, alcuni soggetti che abusano di alcool rientreranno nella categoria di chi ne assume occasionalmente sia all'interno del gruppo dei casi (malati di cancro

del cavo orale) sia all'interno del gruppo dei controlli (sani). L'errore di classificazione del fattore di rischio influenza in egual misura sia i casi che i controlli, determinando una sottostima del reale rapporto fra frequenza di malattia negli esposti al fattore di rischio e nei non esposti.

Misclassificazione Differenziale: l'errore di classificazione (Misclassificazione) del fattore di rischio (o della malattia) dipende dalla malattia (o dal fattore di rischio). L'effetto della Misclassificazione Differenziale è imprevedibile, potendo determinare sia una sottostima che una sovrastima della misura di effetto.

Un esempio tipico di Misclassificazione Differenziale è il *Recall Bias in cui i soggetti malati hanno maggior probabilità di ricordare l'esposizione al fattore di rischio*. Nelle persone malate questo fenomeno trova motivazione nella ricerca personale di una causa, che dia spiegazione della propria malattia. Al contrario i sani (non sentendosi malati) possono esaminare meno attentamente la propria storia passata, omettendo quindi l'esposizione al fattore di rischio.

3. I *Performance Bias* sono dovuti a differenze sistematiche nella somministrazione degli interventi, trattamenti o esposizioni (non comparabilità degli effetti estranei).

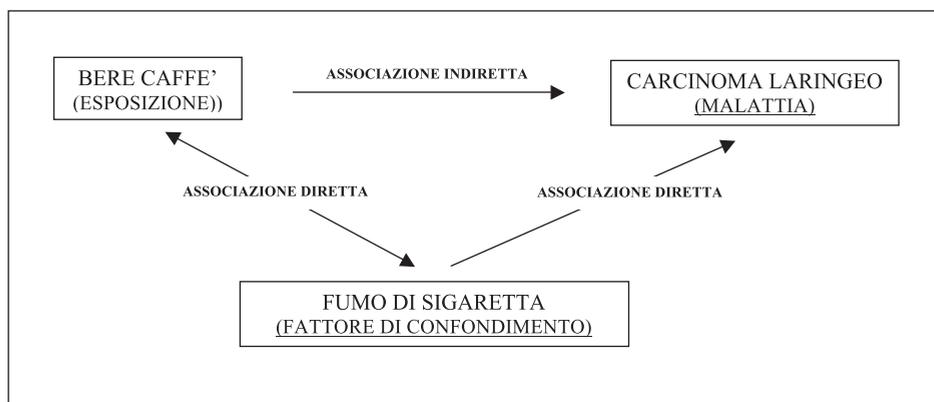
Una corretta programmazione della ricerca è indispensabile per ridurre gli Errori Sistemati che raggiungono valori minimi negli studi randomizzati in doppio cieco grazie alla randomizzazione, alla cecità, all'analisi secondo il principio dell'intenzione al trattamento.

III. L'Associazione osservata fra Esposizione e Malattia può ricondursi a “Fattori di Confondimento”?

Si definisce *Fattore di Confondimento* un *terzo fattore* presente nella popolazione esaminata e che risulta essere *associato sia alla malattia, sia al fattore di rischio* oggetto di studio. Nel caso in cui l'esposizione sia associata alla malattia *solo* perché entrambi sono *associati al Fattore di Confondimento*, si mettono in evidenza *relazioni causa-effetto che in realtà non esistono*. Pertanto l'esposizione rilevata è associata alla malattia solo in modo *indiretto*, in quanto mediata dal Fattore di Confondimento. Due condizioni caratterizzano il Fattore di Confondimento:

- È associato all'esposizione senza esserne una conseguenza
- È associato all'esito indipendentemente dall'esposizione

Riportiamo un esempio di Associazione Statistica Indiretta conseguente al Fattore di Confondimento.



In uno studio che intenda indagare l'associazione fra caffè e Ca laringeo è indispensabile valutare l'effetto di confondimento esercitato dal fumo di sigaretta. I fumatori bevono notoriamente più caffè dei non fumatori, di conseguenza un'eventuale associazione tra caffè e Ca laringeo potrebbe essere causata dal fatto che tra i bevitori di caffè vi siano più fumatori. Il caffè risulta dunque associato solo indirettamente al carcinoma, in quanto correlato al fumo (che rappresenta il reale fattore di rischio per il cancro).

La gestione dei Fattori di Confondimento può essere condotta sia nella *Fase di Progettazione* dello studio, che nell'*Analisi dei Dati*.

1. Nella *Fase di Progettazione* il controllo dei Fattori di Confondimento si realizza mediante:

- la Restrizione in ingresso allo studio: non vengono inclusi nel campione i soggetti con potenziali Fattori di Confondimento.
- la Randomizzazione dei soggetti ai gruppi dello studio: assicura una distribuzione omogenea delle variabili di confondimento conosciute e soprattutto non conosciute tra i gruppi studiati.
- il Matching o Appaiamento in ingresso allo studio: ad ogni soggetto scelto se ne affianca un altro simile per caratteristiche.

2. Nell'*Analisi dei Dati* il controllo dei Fattori di Confondimento si realizza mediante:

- la Stratificazione (per esempio, tenendo fissa l'età si studia l'effetto del sesso o di altre variabili).

- i Modelli Statistici: la tecnica più utilizzata è l'analisi multivariata che consente di poter apprezzare (simultaneamente ed indipendentemente) gli effetti di ciascuna delle variabili considerate, depurati dalle reciproche interferenze.

IV. Esiste una relazione di Causalità fra Esposizione e Malattia?

Dopo aver escluso i Fattori di Confondimento, l'effetto di Errori Sistemati e del Caso, è legittimo concludere che l'associazione osservata fra malattia e fattore di rischio abbia un'*origine Causale*?

La risposta è inequivocabilmente no. Anche in presenza di una programmazione corretta, la ricerca di una relazione causa-effetto è il risultato di uno *studio osservazionale* in cui l'*inferenza Causale fra esposizione e malattia è di per sé stessa debole* per l'impossibilità intrinseca di controllare tutti i fattori di confondimento. Per questa ragione nel 1971 il Ministero della Sanità degli Stati Uniti emise dei canoni logici definiti *Criteri di Causalità* (già pubblicati nel 1965 da Bradford Hill) al fine di distinguere associazioni causali da associazioni non causali.

1. Forza o Intensità dell'Associazione Statistica: se la Forza dell'Associazione Statistica risulta particolarmente elevata, allora difficilmente la relazione esposizione-malattia è giustificabile alla luce di bias o variabili di confondimento. In alcune aree dell'India e dell'Asia in cui è peculiarmente diffusa l'abitudine di masticare tabacco o noce di Betel, l'incidenza di carcinoma del cavo orale è 3 volte più elevata che negli Stati Uniti: è improbabile che un tale risultato sia imputabile a variabili di confondimento e se così fosse dovrebbero essere facilmente identificabili.
2. Sequenza Temporale corretta: se il fattore di rischio interviene causalmente sulla malattia, è essenziale che l'esposizione preceda temporalmente l'effetto che ne deriva (tale criterio è facilmente soddisfatto negli studi sperimentali ed in quelli di coorte).
3. Plausibilità Biologica: gli effetti del fattore di rischio devono essere possibili biologicamente, cioè riconducibili a meccanismi biologici già noti e/o ben definiti.
4. Riproducibilità o Concordanza: riproducendo il medesimo studio in popolazioni e circostanza differenti, i risultati ottenuti devono essere tra loro simili e coerenti.
5. Relazione Dose-Risposta: incrementando la Dose di esposizione al fattore di rischio deve aumentare l'incidenza della malattia.
6. Reversibilità del Rischio: la rimozione dell'esposizione deve implicare una riduzione dell'incidenza della malattia.

7. Specificità: se un fattore di rischio è correlato ad una sola malattia, allora l'interpretazione causale risulta maggiormente sostenibile. Poiché la condizione di Specificità è raramente riproducibile nella realtà, questo criterio non viene ritenuto essenziale per rifiutare una relazione di causa-effetto.
8. Coerenza: la sospetta relazione di causa-effetto non deve entrare in conflitto con la storia naturale della malattia. Ad esempio, le differenze di sesso nella mortalità dovuta al Ca laringeo correlano bene con le differenze abitudinarie di consumo di tabacco ed alcool fra i due sessi.

Conclusioni

La complessità nel soddisfare i Criteri di Causalità in concomitanza all'effetto di Relazioni Causali, Errori Sistemati e Fattori di Confondimento rende rara l'evenienza in cui un singolo studio sia capace di provare da solo la relazione causale fra esposizione e malattia. Epicentro della buona riuscita di uno studio scientifico è il tipo di ricerca che viene elaborato: i trials clinici randomizzati ne rappresentano certamente il modello migliore da riprodurre, anche se spesso inficiati da ragioni di tipo etico che ne limitano la realizzazione. Differenti studi aventi il medesimo quesito di relazione causale, possono essere confrontati tra di loro mediante specifiche tecniche statistiche quali ad esempio le *ri-analisi* e le *meta-analisi*.

La tabella seguente riassume le caratteristiche principali dei diversi disegni di studio, alla luce dei più frequenti bias e variabili di confondimento.

Probabilità di	Studio Ecologico	Studio Trasversale	Studio Caso-Controllo	Studio di Coorte	Trial Randomizzato
<i>Bias di Selezione degli inclusi</i>	/	+++	++++	++	++
<i>Recall Bias</i>	/	++++	++++	++	++
<i>Variabili di Confondimento</i>	++++	+++	+++	++	+

/ Non Applicabile ; ++++ Elevata; +++ Media; ++ Bassa; + Molto Bassa

Bibliografia

1. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. First edition. Oxford: Oxford University Press; 1992.
2. CIOMS (Council for International Organizations of Medical Sciences) and WHO (World Health Organization). International ethical guidelines for biomedical research involving human subjects. Geneva: CIOMS; 1993.
3. Glantz SA. Statistiche per discipline bio-mediche. Terza edizione. Milano: McGraw-Hill Libri Italia s.r.l.; 1994.
4. Grassi M. Statistica in medicina: un approccio basato sulla verosimiglianza. Prima edizione. Milano: McGraw-Hill Libri Italia s.r.l.; 1994.
5. Gupta PC, Mehta FS, Pindborg JJ, et al. Intervention study for primary prevention of oral cancer among 36000 Indian tobacco users. Lancet 1986;1:1235-9.
6. Hill AB. The environment and disease: association or causation? Proceed Roy Soc Medicine 1965;58:295-300.
7. IARC (International Agency for Research on Cancer). IARC monographs on the evaluation of carcinogenic risks to humans. Vol. 38. Tobacco smoking. Lyon: IARC; 1986.
8. Last JM. A dictionary of epidemiology. Third edition. Oxford: Oxford University Press; 1995.
9. Munoz N, Day NE. Cancer epidemiology and prevention. Second edition. Oxford: Oxford University Press; 1996.
10. Sottocornola F. Elementi di metodologia epidemiologica. Prima edizione. Pavia: La Goliardica Pavese s.r.l.; 2003.

FATTORI PROGNOSTICI, INTERPRETAZIONE DEI DATI ONCOLOGICI E LORO ANALISI

F. Chiesa, L. De Benedetto

I fattori prognostici sono dati clinici o strumentali che forniscono informazioni sul decorso e sulla prognosi di una specifica neoplasia e sulla possibile risposta alle diverse opzioni terapeutiche. Questi ultimi vengono più correttamente definiti predittivi. In questo capitolo faremo riferimento soltanto ai fattori e ai markers prognostici.

Il fattore prognostico ideale dovrebbe essere testato mediante test standardizzati, di facile esecuzione e poco costosi. Esso dovrebbe fornire informazioni affidabili sulla evolutività di ogni singola neoplasia allo scopo di programmare le strategie terapeutiche e di controllo ottimali.

Ad oggi questo ideale elemento non esiste: abbiamo la possibilità di acquisire significative informazioni prognostiche da fattori clinici, patologici e molecolari; esiste tuttavia notevole confusione sul loro effettivo significato (ovvero sulla loro affidabilità) e sulle possibili applicazioni in ambito clinico.

Un frequente motivo di confusione nella valutazione delle diverse casistiche e degli studi clinici è dovuto all'uso spesso ambiguo di termini come: fattore di rischio, prognostico, predittivo, marker, biomarker, intermediate-end-point. Infatti talora questi termini vengono considerati come sinonimi, mentre altre volte vengono utilizzati per sottolineare peculiari caratteristiche. Ad esempio all'inizio di questo capitolo abbiamo sottolineato la differente valenza del termine prognostico (indicativo di outcome, ovvero delle evolutività di una neoplasia) e di quello predittivo (suggestivo di responsività della neoplasia ad una determinata terapia). Questi due termini in letteratura o nella pratica clinica vengono molte volte usati con lo stesso significato, in genere per quantizzare la probabilità di ricadute loco regionali o di progressione della neoplasia. Chiarito il significato di fattori prognostici per standardizzare il loro uso nella pratica clinica possiamo distinguerli in tre gruppi:

- **relativi al paziente:** dati demografici, quali età e sesso, performance status, familiarità e co-morbidità;
- **propri della neoplasia:** sede e dimensioni della neoplasia, numero e livello dei linfonodi metastatici, istopatologia, biologia molecolare;
- **legati alla terapia:** la scelta terapeutica ed eventuali errori di pianificazione della stessa, possono condizionare pesantemente la prognosi, l'evoluzione e la possibilità di recupero.

La significatività dei fattori prognostici dipende dal rigore con cui sono stati disegnati e condotti gli studi clinici, che devono rispondere a ben definiti standard quali: a) casistiche adeguate: almeno 100 pazienti per studi retrospettivi e 50 per quelli prospettici; b) follow-up minimo di due anni per il controllo loco-regionale della malattia e cinque anni per la sopravvivenza; c) analisi statistica multivariata dei risultati.

Analizzando i vari lavori secondo questi criteri possiamo valutare l'affidabilità di ogni singolo fattore studiato:

- **Fattori sicuramente significativi:** sono da tutti gli Autori riconosciuti come affidabili e predittivi dell'evoluitività della neoplasia. Sono quelli che devono sempre essere verificati e riportati in cartella clinica rappresentando lo standard clinico minimo richiesto.
- **Fattori probabilmente significativi:** non vi è consenso unanime sulla loro significatività. Sono tecnicamente eseguibili di routine in ogni laboratorio, meritano ulteriori valutazioni mediante studi clinici.
- **Fattori sperimentali:** non ancora standardizzati, anche se clinicamente promettenti, sono valutabili con test complessi o di non agevole accesso. Non sono utilizzati ancora nella pratica clinica, ma solo in studi di fase I o II in centri specialistici.

Questo capitolo si propone di indicare quali possono essere considerati i fattori prognostici più affidabili oggi disponibili in oncologia cervicofacciale, ma soprattutto di fornire utili indicazioni nella interpretazione dei numerosi lavori che vengono sempre più frequentemente pubblicati su questo argomento. Questo permette di valutare criticamente quali markers possono fornire informazioni utili alla programmazione terapeutica e quali invece non rivestono un significato clinico, ma sono piuttosto argomento di studio e di approfondimento. Questa consapevolezza consente di evitare di utilizzare routinariamente nella pratica clinica tests costosi dal significato prognostico incerto o non accertato, consentendo quindi di ridurre i costi diagnostici e di evitare al paziente inutili tensioni psicologiche legate alla incerta interpretazione dei loro risultati.

Fattori correlati al paziente

ETÀ

Il carcinoma squamoso della testa-collo insorge spesso tra la quinta e sesta decade di vita. L'età di per sé non rappresenta un fattore prognostico negativo, che è invece spesso correlato alle comorbidità, che sono più frequenti o più serie nel paziente anziano.^{1,2}

Nelle neoplasie delle ghiandole salivari è descritta una migliore prognosi per i pazienti con meno di 60 anni, questo è stato attribuito al fatto che nei giovani siano più frequenti neoplasie a bassa malignità^{3, 4}.

Anche nei tumori tiroidei differenziati, l'età alla diagnosi ha un ruolo prognosticamente rilevante, tanto che molti Autori ed il TNM prevedono due differenti classificazioni nei pazienti di età inferiore o superiore a 45 anni^{5, 6}.

SESSO

È nota la prevalenza nel sesso maschile del carcinoma squamoso, molti Autori però correlano questo dato ad un' aumentata esposizione a fattori di rischio, come l'abuso di alcool ed il tabagismo. Per questo non vi sono dati concordanti circa l'impatto del fattore sesso sulla sopravvivenza o sulla insorgenza di recidive nelle analisi multivariate pubblicate^{7, 8}. Nel tumore differenziato della tiroide, non è più considerato un fattore influente^{5, 6}.

FUMO ED ALCOOL

Il fumo di sigaretta ed il consumo di bevande alcoliche sono certamente i fattori cancerogeni più importanti per quanto riguarda lo sviluppo del carcinoma squamocellulare del distretto testa-collo⁹⁻¹⁵. Secondo alcuni Autori¹⁰ queste abitudini (in particolare l'abuso di tabacco) condizionano la risposta alla radioterapia di neoplasie del cavo orale. Bundgaard¹¹ invece ha dimostrato una relazione diretta tra consumo di alcool e di tabacco e sopravvivenza a 5 anni all'analisi univariata (54% nei non bevitori contro il 39% dei bevitori: $p=0.03$), mentre l'analisi multivariata ha confermato la significatività solo per il fumo ($p=0.046$). Secondo Mak-Kregar ed Harari^{12, 13} il tabagismo influenza pesantemente la sopravvivenza a 5 anni (68% vs 32%: $p=0.0019$). Nei tumori dell'orofaringe l'alcool agirebbe come solvente dei prodotti di combustione del tabacco determinando un insidioso sinergismo.

Calhoun¹⁴ su 727 casi di tumore delle VADS ha evidenziato un aumentato rischio di metastasi polmonari, in proporzione al consumo di alcool e tabacco.

Nei pazienti anziani invece, Chin¹⁵ non ha riscontrato riduzione significativa delle sopravvivenze dei soggetti appartenenti alle due categorie.

STATO NUTRIZIONALE

La malnutrizione spesso associata ad abitudini di vita voluttuarie come il consumo di alcool oltre ad essere un fattore di rischio nella cancerogenesi è causa di una più scarsa attitudine a sopportare trattamenti terapeutici aggressivi. Alcuni studi recenti hanno cercato di dimostrare

questa correlazione come ad esempio il lavoro di Rabinovitch¹⁶ che ha misurato l'impatto del supporto nutrizionale sui risultati oncologici nei pazienti con HNSCC localmente avanzato trattati con radioterapia.

FAMILIARITÀ

Nei pazienti affetti da HNSCC con familiarità positiva, non esistono dati in letteratura che ne definiscano una miglior o peggior prognosi rispetto ai pazienti senza familiarità.

Nel caso del tumore midollare della tiroide in pazienti con MEN2A o FMTC sembra che la migliore prognosi sia essenzialmente dovuta a diagnosi precoce grazie a studi di screening per identificazione del gene RET e successiva tiroidectomia profilattica eseguita tanto più precocemente quanto più alto il rischio dedotto dall'analisi genomica¹⁷.

FATTORI ETNICO-RAZZIALI

Il tumore follicolare e midollare della tiroide dai pochi dati disponibili in letteratura sembrano avere tassi di sopravvivenza peggiori nei pazienti di origine afro-americana rispetto a quelli di altre etnie¹⁸.

COMORBIDITÀ

Il fumo di sigaretta ed il consumo elevato di alcool sono come sappiamo causa della cancerogenesi dell'HNSCC, ma le due sostanze sono anche associate ad altre comorbidity sistemiche polmonari, cardiache e metaboliche. Diversi studi hanno dimostrato che queste comorbidity hanno un impatto significativamente negativo sulla sopravvivenza dei pazienti con HNSCC. Secondo Chin-Ting et al¹⁹ la sopravvivenza a 3 anni passerebbe da 25.9% a 3.5% in pazienti con stadio III e IV trattati con radioterapia e/o chemio-radioterapia con gravi comorbidity sistemiche. Secondo Hall et al, su un campione più numeroso di 655 pazienti, la sopravvivenza a 5 anni passerebbe dal 59% nel sottogruppo senza comorbidity al 23% in quello con patologie sistemiche concomitanti, in particolare nella sua esperienza la pneumopatia cronica e le epatopatie moderate o severe²⁰.

Fattori correlati alla neoplasia

SEDE ED ESTENSIONE DELLA NEOPLASIA

Nel distretto cervico-facciale la sede posteriore ha in genere una prognosi meno favorevole rispetto a quella anteriore, verosimilmente perché in quella sede la rete linfatica sottomucosa è più fitta e favorisce la diffusione locale e regionale. L'estensione locale della neoplasia ed in particolare l'invasione ossea sono fattori prognostici sfavorevoli. La presenza di meta-

stasi linfonodali, soprattutto se multiple o con rottura capsulare dimezza la probabilità di sopravvivenza rispetto ai pazienti classificati N0.^{7-9,12}

ESPRESSIONE DEL RECETTORE EPITHELIAL GROWTH FACTOR (EGFR)
Si tratta di un marcatore studiato in molte neoplasie tra cui quelle della testa-collo. Chung²¹ ha evidenziato che i pazienti con scarsa o nulla espressione dell'EGFR hanno una sopravvivenza peggiore rispetto ai pazienti che lo esprimono.

Recentemente è stato segnalato che l'espressione dell'EGFR sembra essere un indice di sensibilità al cetuximab, chemioterapico che per questo motivo viene associato alla radioterapia in molte neoplasie del distretto cervico-facciale²².

INFEZIONE DA HUMAN PAPILLOMA VIRUS (HPV)

Il gruppo dell'MD Anderson Cancer Center di Huston, USA, ha recentemente pubblicato una metanalisi di 34 studi sulla correlazione tra infezione da HPV e HNSCC per un totale di 5681 pazienti. Questa correlazione è stata dimostrata nel 41% dei carcinomi dell'orofaringe, contro una media del 22% di tutte le altre sottosedi del distretto testa-collo. Questi studi hanno dimostrato che il rischio di sviluppare un tumore squamoso della testa-collo nei soggetti positivi all'HPV-16 è di 4,4 volte maggiore rispetto agli HPV-16 negativi. Questo rischio per i fumatori e per i bevitori è rispettivamente di 3.53 e 4.04 volte.

I risultati di numerosi studi clinici hanno documentato una risposta alla radio ed alla radio-chemioterapia nettamente migliore nei pazienti con carcinomi HPV correlati rispetto a quelli non-HPV correlati²³.

Interpretazione dei dati oncologici e loro analisi

L'interpretazione dei dati oncologici riportati in un lavoro scientifico, dovrebbe essere condotta con uno spirito critico, analizzando significatività ed affidabilità dello studio in esame. In questo modo si possono evidenziare eventuali bias insiti nella impostazione dello studio in esame. Segnaliamo alcuni fra i dati più significativi da considerare:

Il **tipo di studio**: per l'analisi dei fattori prognostici il più indicato è lo studio di coorte, o caso controllo, prospettico; al contrario uno studio retrospettivo è più esposto ad errori (bias) con conseguente minor forza di inferenza clinica.

Il **campionamento dei pazienti**, deve essere rappresentativo della popolazione. Uno dei bias nell'arruolamento dei pazienti è dovuto al tipo di Centro dove viene eseguito l'arruolamento. I Centri di livello superiore

facilmente trattano casi più complessi che possono quindi presentare un maggior numero di eventi avversi rispetto ad altre casistiche seguite in centri meno specializzati.

Il **follow-up** deve essere adeguato e completo, devono sempre essere ricercati i motivi per cui i pazienti vengono persi di vista (spesso i pazienti anziani non si presentano ai controlli oncologici per decadimento delle condizioni generali, per ricovero in strutture ospedaliere per cronici); alcuni pazienti si rivolgono invece ad altre strutture perché nel frattempo si sono trasferiti in altre città. Anche la causa di morte deve sempre essere sempre indagata, per cercare di chiarire se sia correlata alla neoplasia o ad altre cause. Un lavoro con un elevato numero di pazienti persi di vista (superiore al 15%) perde gran parte della sua significatività.

La **valutazione degli esiti** (outcomes) deve essere eseguita sempre con criteri obiettivi e rigorosi.

Un fattore prognostico può essere alterato dalla presenza di altri fattori concomitanti (comorbidità e trattamenti in atto) confondenti. Per definire il reale peso del fattore prognostico in esame occorre attuare l'**aggiustamento degli altri fattori prognostici**, per il quale è necessaria una competenza statistica.

ESEMPIO

In questa sede abbiamo preferito scegliere e indicare alla valutazione dei lettori una review che ci sembra particolarmente utile perché fornisce gli elementi indispensabili non solo alla interpretazione corretta dei risultati di lavori sui fattori prognostici, ma anche indica le regole da seguire nella impostazione di uno studio clinico (sia retrospettivo che prospettico). Il lavoro preso in considerazione è il seguente²⁴: Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single study to meta-analysis, and beyond. *BJC* 2009; 100:1219-29. Il testo integrale dell'articolo può essere scaricato gratuitamente dal sito: <http://www.ncbi.nlm.nih.gov/pubmed>.

Invitiamo a considerare alcuni punti critici evidenziati in questa review:

- a. i lavori sui fattori prognostici disponibili in letteratura sono moltissimi, ma pochi di questi fattori hanno un ruolo significativo nella pratica clinica quotidiana;
- b. spesso vengono studiati fattori o markers giudicati promettenti o per lo meno meritevoli di ulteriori approfondimenti (studi su casistiche più numerose, standardizzazione delle metodiche,..). Tuttavia nella maggioranza dei casi questi lavori non hanno seguito: Brundage²⁵ ha analizzato oltre 100 markers potenzialmente prognostici nei carcinomi polmonari non a piccole cellule, la media dei lavori pubblicati per ognuno di questi fattori era 1. Nessuno di queste

ricerche ha quindi avuto seguito. Pertanto attenzione a valutare la significatività di un lavoro, anche se ben scritto e senza apparenti bias o fattori di confondimento: conviene fare una ricerca in letteratura per vedere se gli stessi Autori o altri gruppi lo hanno valutato e validato.

- c. Disegno dello studio: spesso non è riportato il rationale della scelta di un determinato marcatore; solo una piccola parte di questi studi è prospettica: secondo Kyzas solo il 20% su 331 studi considerati²⁶. Se il vantaggio di uno studio retrospettivo sta nella possibilità di correlare l'outcome dei pazienti con il fattore considerato, lo svantaggio sta nel fatto che questi studi spesso non hanno un rationale solido ed il disegno dello studio è spesso approssimativo; non sempre sono segnalati con precisione i criteri di inclusione; il campione in esame frequentemente non è ben definito; i fattori considerati non hanno procedure diagnostiche standardizzate; i dati spesso sono incompleti ed il follow-up è irregolare.
- d. La valutazione statistica deve contenere sia l'analisi univariata, che quella multivariata ed è consigliabile che sia basata su modelli predefiniti²⁷.

In conclusione ci sembra indispensabile che ogni persona realmente interessata all'approfondimento di questi argomenti impari ad analizzare le basi teoriche dello studio: solo una ricerca costruita su un disegno meditato, giustificata da un rationale solido e condotta con rigore metodologico deve essere considerata affidabile.

Bibliografia

1. Boffetta P, Merletti F, Magnani C, Terracini B. A population-based study of prognostic factors in oral and oropharyngeal cancer. *Eur J Cancer B Oral Oncol* 1994; 30: 369-73.
2. Martin-Granizo R, Rodriguez-Campo F, Naval L, Diaz Gonzalez FJ. Squamous cell carcinoma of the oral cavity in patients younger than 40 years. *Otolaryngol Head Neck Surg* 1997; 117: 268-75.
3. Frankenthaler RA, Luna MA, Lee SS et al. Prognostic variables in parotid gland cancer. *Arch Otolaryngol Head Neck Surg* 1991; 117: 1251-6.
4. Frankenthaler R, Byers RM, Luna et al. Predicting occult lymph node metastasis in parotid cancer: *Arch Otolaryngol Head Neck Surg* 1993; 119: 517-520.
5. Shaha AR. Prognostic factors in papillary thyroid carcinoma and implications of large metastasis. *Surgery* 2004; 135: 237-9.
6. Shaha AR, Shah JP, Loree TR. Low-risk differentiated thyroid cancer: the need for selective treatment. *Ann Sur Oncol* 1997; 4: 328-33.
7. Gehanno P, Enaud M, Barry B et al. Cancer in the tongue in women. A propos of 40 cases. *Ann Otolaryngol Chir Cervicofac* 1994; 111: 265-9.
8. Urist MM, O'Brien CJ, Soong SJ et al. Squamous cell carcinoma of the buccal mucosa: analysis of prognostic factors. *Am J Surg* 1987; 154: 411-41.
9. Oliver AJ, Helfrick JF, Gard D. Primary oral squamous cell carcinoma: a review of 92 cases. *J Oral Maxillofacial Surg* 1996; 54: 949-54.
10. Browman GP, Wong G, Hodson I et al. Influence of cigarette smoking on the efficacy of radiation therapy in head and neck cancer. *New Engl J Med* 1993; 328: 159-63.
11. Bundgaard T, Bentzen SM, Wildt J. The prognostic effect of tobacco and alcohol consumption in oral squamous cell carcinoma. *Eur J Cancer B Oral Oncol* 1994; 30: 323-8.
12. Mak Kregar S, Hilger FJ, Baris G et al. Carcinoma of the tonsillar region: comparison of two staging systems and analysis of prognostic factors. *Laryngoscope* 1990; 100: 634-8.
13. Harari P, Kinsella J. Advances in radiation therapy for head and neck cancer. *Curr Opin Oncol* 1995; 7: 248-54.
14. Calhoun KH, Fulmer P, Weiss R, Hokanson JA. Distant metastases from head neck squamous cell carcinomas. *Laryngoscope* 1994; 104: 1199-205.
15. Chin R, Fischer RJ, Smee RI, Barton MB. Oropharyngeal cancer in elderly. *Int J Radiat Oncol Biol Phys* 1995; 32:1007-16.
16. Rabinovitch R, Grant B, Berkey BA, Raben D. Impact of nutrition support on treatment outcome in patients with locally advanced head and neck squamous cell cancer treated with definitive radiotherapy: a secondary analysis of RTOG trial 90-03. *Head Neck* 2006; 28: 287-96.
17. Schreinemakers JM, Vriens MR, Val KGD et al. Factors predicting outcome of total thyroidectomy in young patients with multiple endocrine neoplasia type 2: a nationwide long-term follow-up study. *World J Surg* 2010; 34: 852-860.

18. Gilliland FD, Hunt WC, Morris DM, Key CR. Prognostic factors for thyroid carcinoma. *Cancer* 1997; 79: 567-73.
19. Chin-Ting Liu, Tai-Jan Chiu, Tai-Lin Huang et al. Impact of comorbidity on serviva for locally advanced head and nech cancer patients treated by radioterapy or radiotherapy plus chemotherapy. *Chang Gung Med J* 2010; 33: 283-290.
20. Hall SF, Groome PA, Rothwell D. The impact of comorbidity on the survival of patients with squamous cell carcinoma of the head and neck. *Head Neck* 2000; 22: 317-22.
21. Chung CH, Ely K, McGavran L, Varella-Garcia M et al. Increased epidermal growth factor receptor gene copy number is associated with poor prognosis in head and neck squamous cell carcinomas. *J Clin Oncol* 2006; 24: 4170-6.
22. Cripps C, Winqvist E, Devries MC et al. Epidermal growth factor receptor targeted therapy in stages III and IV head and neck cancer. *Curr Oncol* 2010; 17: 37-48.
23. Dayyani F, Etzel CJ, Liu M et al. Meta-analysis of the impact of human papilloma virus (HPV) on cancer risk and overall serviva in head and neck squamous cell carcinomas (HNSCC). *Head Neck Oncol* 2010; 2: 15.
24. Riley RD, Sauerbrei W, Altman DG. Prognostic markers in cancer: the evolution of evidence from single study to meta-analysis, and beyond. *BJC* 2009; 100:1219-29.
25. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. *Chest* 2002; 122:1037-57.
26. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Quality of reporting of cancer prognostic markers studies: association with reported prognostic effect. *J Natl Cancer Inst* 2007; 99:236-43.
27. Boracchi P, Biganzoli E. Markers of prognosis and response to treatment: ready for clinical use in oncology? A biostatistician's point of view. *Int J Biol Markersn* 2003; 18:65-9.

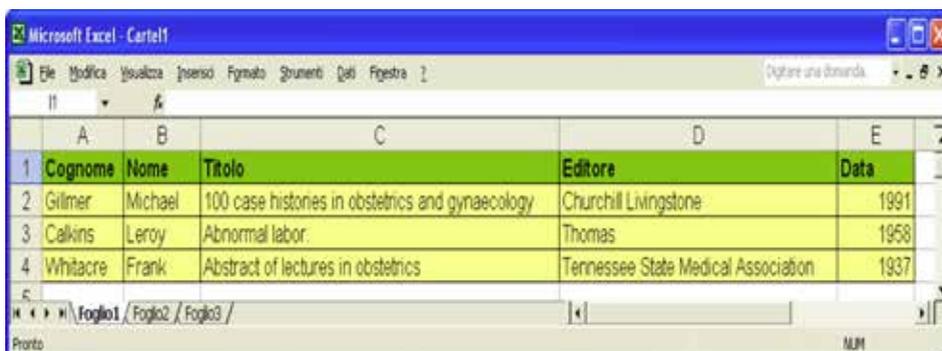
VALIDITÀ, LIMITI E RISCHI DELLA RICERCA DATI TRAMITE BANCHE ELETTRONICHE NELLA PRATICA CLINICA

A. Camaioni , C. Viti , M. Simone

Le banche elettroniche

I **database** (base di dati, banca dati, base dati, archivio di dati), nella definizione più semplice, sono degli archivi di dati strutturati, più spesso in tabelle divise per argomento. Ogni **tabella** è formata da diverse **celle** (elementi) contenenti dati di vario tipo (numeri interi, decimali, a virgola mobile, stringhe di caratteri, etc.) che sono le unità minime della tabella stessa; il **record**, identificabile con una riga della tabella, è una struttura eterogenea di elementi che insieme definiscono le caratteristiche del record stesso e costituisce l'unità base del database. Al contrario le colonne della tabella sono definite **campi**, ossia un insieme di elementi omogenei tra loro, identificati da un'etichetta (tag), che hanno lo scopo di caratterizzare i vari record (vedere fig. 1).

In campo medico i database sono prevalentemente di tipo bibliografico (l'esempio più famoso è PubMed/MEDLINE). In questo tipo di database i record sono rappresentati dalle citazioni bibliografiche (che derivano dallo spoglio di articoli delle riviste prese in considerazione dal database) ed ognuna di queste singole citazioni è organizzata in campi (Autore, titolo, rivista, etc.; in MEDLINE ogni citazione ha circa 40 campi di caratterizzazione).



The image shows a screenshot of a Microsoft Excel spreadsheet titled 'Cartell'. The spreadsheet contains a table with 5 columns and 4 rows of data. The columns are labeled 'Cognome', 'Nome', 'Titolo', 'Editore', and 'Data'. The rows contain the following data:

	A	B	C	D	E
1	Cognome	Nome	Titolo	Editore	Data
2	Grimer	Michael	100 case histories in obstetrics and gynaecology	Churchill Livingstone	1991
3	Calkins	Leroy	Abnormal labor.	Thomas	1958
4	Whitacre	Frank	Abstract of lectures in obstetrics	Tennessee State Medical Association	1937

Fig. 1: In questo esempio di semplice database i record sono le varie righe caratterizzate da più celle (elementi) ognuna appartenente ad un campo (colonna) descritto da un'etichetta (cognome, nome, titolo, etc.).

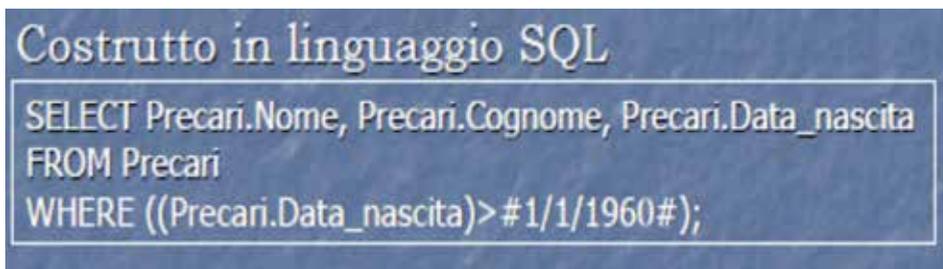


Fig. 2: Esempio su come interrogare un database contenente i dati dei dipendenti di un ospedale, impostando una query in linguaggio SQL, per sapere quanti e quali sono i dipendenti precari nati dal 1960 in poi.

Il tipo di database più frequentemente utilizzato è quello cosiddetto “relazionale”, ossia un database in cui le singole tabelle sono in relazione fra loro secondo una modalità detta *univoca*:

- a. in cui ad ogni record di un determinato campo in una tabella A corrisponde uno ed un solo record della tabella B;
- b. oppure ad ogni record di un determinato campo nella tabella A corrisponde un solo record nella tabella B ma più record della tabella A possono corrispondere allo stesso record della tabella B;

o una modalità detta *non univoca*:

1. in cui ad ogni record di un determinato campo nella tabella A possono corrispondere uno o più record di un determinato campo nella tabella B, e viceversa.

Per interrogare un database con questa incredibile capacità di interrelazione fra i dati fu creato un linguaggio adatto: il SQL (Structured Query Language). Lo svantaggio di questo linguaggio è che, per i non addetti ai lavori, voler interrogare un database in questo modo risultava alquanto ostico (vedi fig. 2).

Per questo motivo attualmente i software di gestione dei database hanno delle maschere grafiche attraverso cui poter impostare delle query in modo più semplice ed intuitivo. L'unico passaggio, quindi, che in sostanza è passibile di errori, ingenerando risultati non soddisfacenti dal punto di vista della ricerca, è proprio *l'impostazione della query*.

I requisiti fondamentali per condurre una buona ricerca bibliografica sono:

1. conoscere il funzionamento del motore di ricerca;
2. applicare strategie di ricerca che permettano di individuare gli articoli rilevanti per la soluzione del quesito posto.

Il funzionamento del motore di ricerca

Prima di parlare della strategia di ricerca è giusto fare un accenno al funzionamento dei database biomedici e dei loro motori di ricerca. È molto importante, infatti, conoscere il funzionamento di questi ultimi in modo da adattare al meglio la strategia al database utilizzato. Bisogna, però, ricordare che esistono numerosi motori di ricerca di banche dati biomediche ed è praticamente impossibile sperare di conoscerli tutti. Per questo motivo cercheremo di soffermarci sul principale tra questi (le cui caratteristiche salienti sono comunque simili in tutti): il PubMed/MEDLINE.

Prodotto dalla National Library of Medicine (NLM), è la banca dati più diffusa ed attualmente indicizza circa 5000 riviste con una raccolta di quasi 18.000.000 di citazioni bibliografiche. L'accesso al MEDLINE, inizialmente possibile solo ad istituzioni dotate di costosi collegamenti con la NLM, si è progressivamente diffuso grazie all'avvento di internet. Per i contenuti clinici, la diffusione e, soprattutto, la disponibilità gratuita su Internet, il MEDLINE costituisce la banca dati di riferimento⁶.

I limiti principali di questo database derivano dal fatto che, sebbene indicizzi una gran quantità di riviste del settore, in realtà queste sono appena il 50% di tutte quelle mondiali, che la gran parte di queste riviste sono esclusivamente in lingua inglese e che indicizza quasi esclusivamente riviste di articoli scientifici raccogliendo poche linee guida e revisioni sistematiche.

Nel database MEDLINE ogni citazione di articolo (record) contiene oltre 40 campi di caratterizzazione tra cui quello più importante per la nostra ricerca è quello dei **MeSH**.

Il MeSH (Medical Subject Headings) Thesaurus è un vocabolario controllato, aggiornato annualmente, che nella versione del 2010 contiene oltre 23.000 termini organizzati gerarchicamente in una struttura ad albero con 16 gruppi principali. Seguendo un protocollo dettagliato, gli operatori della NLM assegnano ad ogni articolo un certo numero di termini MeSH (massimo fino a 15) che ne descrivono i contenuti ed alcuni termini speciali: *subheading* (servono a caratterizzare meglio i termini MeSH), *major topic* (sono i termini MeSH che descrivono gli aspetti principali dell'articolo e vengono segnalati con un asterisco), *check tags* (caratterizzano alcuni aspetti dell'articolo indicizzato come il disegno dello studio, età e sesso dei soggetti inclusi, area geografica di conduzione, etc.) e *publication type* (descrive il tipo di pubblicazione: trial clinico, articolo di rivista, review, etc.).

In PubMed la query al database può essere impostata utilizzando la cosiddetta "ricerca libera" oppure quella tramite "indicatori MeSH".

LA RICERCA LIBERA

Ognuno di noi è portato, istintivamente, ad effettuare una ricerca in un database inserendo nell'apposito box le parole che ritiene chiave per l'argomento oggetto della ricerca, attingendo al linguaggio naturale.

In questo tipo di ricerca, non esiste alcun procedimento specifico d'indicizzazione, ma il computer riconosce stringhe di testo in uno o più campi del database, restituendoci come risultati tutti i record in cui il riconoscimento ha dato esito positivo.

Oltre agli ostacoli legati alla struttura fisica delle parole (soprattutto derivanti da una non corretta sintassi), il *limite principale* del testo libero consegue al fatto che i termini contenuti nel titolo/abstract di un articolo non fanno parte di un vocabolario controllato, per cui non c'è alcuna garanzia che questi siano gli stessi che useranno gli utenti del MEDLINE per effettuare le ricerche.

Ad esempio, utilizzando la parola *hyperlipidemia* come testo libero, ritroveremo tutte le citazioni dove nel titolo/abstract è presente *hyperlipidemia*, ma non altri termini come ad esempio *hypercolesterolemia*, *hypertriglyceridemia*, etc.; inoltre l'utilizzo di parole tradotte nell'inglese europeo ci faranno perdere tutti quegli articoli in cui la stessa parola è scritta in inglese americano (potrebbero essere differenti) o ad esempio tutti quegli articoli in cui la stessa parola è rappresentata da sinonimi o parafrasi. Infine questo tipo di ricerca sarà molto poco specifica in quanto ci restituirà anche tutti quegli articoli in cui la parola cercata è solo citata, ma non riveste importanza per l'argomento trattato dall'articolo stesso (es. cerco le complicanze dei tumori della mammella e mi torna un articolo in cui si parla di una donna con tumore del retto in periodo di allattamento). In poche parole spesso la ricerca libera ci fornisce risultati molto numerosi, ma poco rilevanti, con notevole dispendio di energie e tempo per effettuare il filtraggio "manuale" dei risultati.

LA RICERCA CON INDICIZZAZIONE MESH

Al contrario la ricerca in PubMed (ma negli altri database il concetto è simile) può essere resa più sensibile e specifica utilizzando gli indicatori MeSH. Diversi studi, infatti, hanno dimostrato che utilizzando il vocabolario controllato si ottengono ricerche più specifiche rispetto all'uso del testo libero¹.

Il principale vantaggio di utilizzare i termini MeSH nella ricerca bibliografica è legato soprattutto al fatto che utilizzando queste "parole controllate" non ci sono rischi di errata sintassi, vengono integrati tutti i sinonimi e le parafrasi (perlomeno quelle semplici) e soprattutto ogni articolo che è stato indicizzato da un operatore con un determinato MeSH si suppone tratti specificamente l'argomento caratterizzato dal MeSH stesso (es. se

un articolo fra i 15 MeSH assegnati ha l'iperlipidemia e l'ipertensione, sarò sicuro che questo articolo tratterà in modo rilevante l'ipertensione e l'iperlipidemia – a questo proposito ricordare i major topics).

Tuttavia, l'indicizzazione umana ha alcuni *limiti*:

- a. la riproducibilità nell'assegnare i termini MeSH è condizionata sia dall'esperienza precedente degli operatori, sia dalla differente interpretazione del protocollo e varia dal 34% per la combinazione MeSH/subheading al 75% per il check tags¹;
- b. nei settori dove il turnover delle conoscenze è estremamente rapido (ad es. AIDS) il vocabolario controllato diventa obsoleto prima dell'aggiornamento annuale;
- c. vengono documentate ripetutamente doppie indicizzazioni.

Le strategie di ricerca

Parliamo, ora, di come applicare strategie di ricerca che portino ad individuare gli articoli più rilevanti, ricordando che queste strategie possono essere impiegate sia nella ricerca libera sia, molto più proficuamente, nella ricerca per indicizzatori MeSH.

Ogni strategia di ricerca possiede due proprietà fondamentali:

- Sensibilità o “recall” → è la capacità di ritrovare le citazioni rilevanti.
- Specificità o “precision” → è la capacità di escludere le citazioni irrilevanti.

Idealmente una strategia perfetta dovrebbe avere una sensibilità del 100% (trovare tutti gli articoli rilevanti) ed una specificità ugualmente del 100% (escludere tutti quelli non rilevanti).

In realtà, a causa di varie limitazioni (es. l'architettura fisica del database, la scarsa riproducibilità dell'indicizzazione umana, ed altre), questo risultato è praticamente impossibile da ottenere. Cercheremo, perciò, di suggerire qualche accorgimento che ci permetta di avvicinarci il più possibile alla situazione ideale.

COME PROCEDERE

A differenza del ricercatore che necessita di sapere “tutto il possibile” su un certo argomento e può dedicare molto tempo alla ricerca bibliografica, al medico è sufficiente, quello che Gardner ha definito “a small representative sample of information”².

In altre parole, poiché la EBM serve al medico per affrontare i quesiti clinici che ogni giorno gli si pongono innanzi, questi non può effettuare ricerche esaustive che poi non avrà tempo di esaminare, ma dovrà porsi

come obiettivo quello di ottenere una breve lista di citazioni (10-15) pertinenti al problema del paziente, metodologicamente corrette e possibilmente pubblicate in riviste reperibili⁵.

Per raggiungere questo obiettivo è necessario⁸:

- definire una strategia di ricerca molto sensibile, per evitare di escludere citazioni rilevanti;
- perfezionarla progressivamente eliminando via via citazioni irrilevanti (rendere la ricerca più specifica);
- filtrare i risultati della ricerca “manualmente”

DEFINIRE LA STRATEGIA DI BASE

Una strategia che più di altre si è andata affermando, probabilmente per la sua semplicità e versatilità rispetto alle varie situazioni cliniche, è la cosiddetta strategia PICO^{3,4,9}.

PICO è l'acronimo di **P**atients group, **I**ntervention/treatment, **C**omparison ed **O**utcome e sta ad indicare le parti in cui scomporre il quesito per effettuare una ricerca il più mirata possibile.

1. Patient group → ossia descrivere le caratteristiche salienti del gruppo di pazienti o del paziente di cui si vuole effettuare la ricerca;
2. Intervention-treatment → puntualizzare quale è l'intervento (trattamento medico, chirurgico, radiologico, fisioterapico, etc.) che si vuole attuare;
3. Comparison → quali sono, se esistono, le alternative principali al trattamento proposto;
4. Outcome → quali sono i risultati attesi (possono essere anche gli eventi avversi, le complicazioni, etc.).

Questa strategia può essere usata indipendentemente sia con la ricerca libera sia con il vocabolario MeSH (si suggerisce di usare sempre, laddove possibile, gli indicizzatori MeSH). Poiché, però, il vocabolario MeSH non è molto intuitivo ed inoltre è molto ampio (più di 23000 voci), è consigliabile usare un convertitore dal linguaggio naturale a quello indicizzato: il PubMed offre in tal senso un eccellente browser, che consente anche di “navigare” all'interno della gerarchia MeSH⁷.

La ricerca mediante testo libero viene generalmente utilizzata quando non esiste il termine MeSH corrispondente (ad esempio quando si cercano nuovi farmaci).

GLI OPERATORI BOOLEANI

Per affinare e rendere più specifica la ricerca ogni parola (sia libera che MeSH) può essere “combinata” con altre tramite i cosiddetti operatori booleani (dall'inventore, il matematico inglese George Boole).

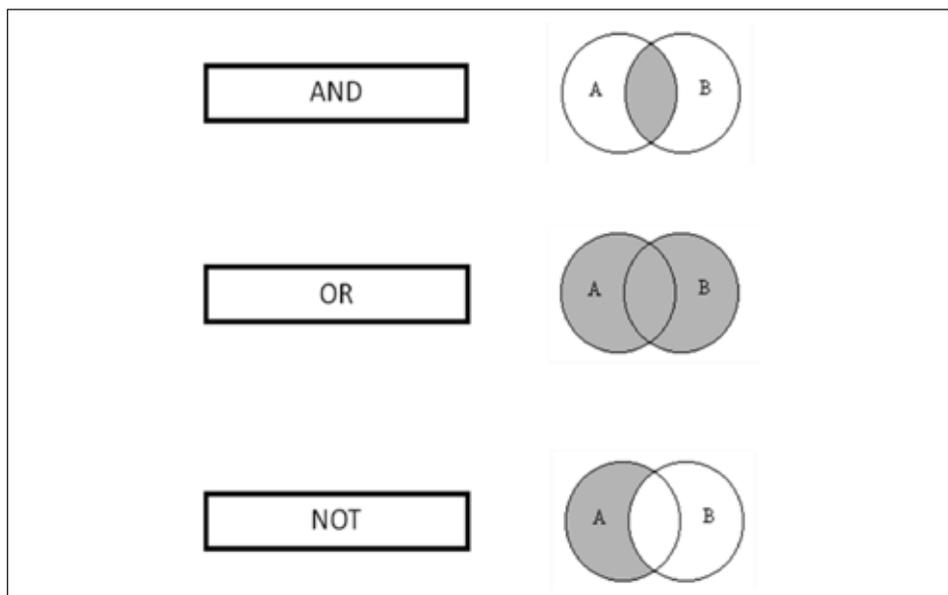


Fig. 3: In questa figura sono rappresentate schematicamente le funzioni degli operatori booleani mediante la rappresentazione grafica ad insiemi.

Sono delle stringhe di testo che il motore di ricerca interpreta come comandi. In Pubmed si possono usare tutti e tre gli operatori (vedi figura 3):

1. AND → è l'operatore detto di *intersezione*; usandolo il motore di ricerca effettuerà "l'intersezione" dei due insiemi composti da tutte le citazioni contenenti la prima parola e da tutte quelle contenenti la seconda, restituendoci come risultato solo l'insieme delle citazioni in cui sono presenti entrambe;
2. OR → operatore detto di *unione*; utilizzando questo operatore il risultato sarà l'insieme delle citazioni dato dalla somma dell'insieme delle citazioni in cui è presente la prima parola (o entrambe) più l'insieme delle citazioni in cui è presente la seconda;
3. NOT → operatore di *complementazione*; usando questo operatore il risultato sarà l'insieme di citazioni che conterranno esclusivamente la prima parola cercata (verranno escluse tutte le citazioni in cui è contenuta la seconda, anche se sono presenti entrambe).

Bisogna ricordare che in Pubmed la ricerca può essere effettuata anche con sintassi complesse, ad esempio mediante l'uso di parentesi, ricordandosi che la stringa viene interpretata sempre da sinistra verso destra. È diverso, per esempio, scrivere larynx OR nose NOT oral (il motore cercherà tutte le citazioni che contengono larynx e nose

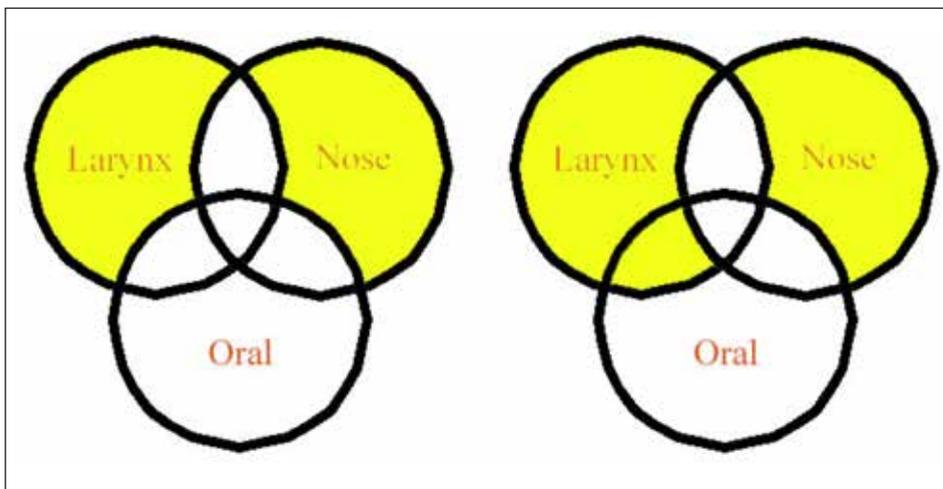


Fig. 4: Gli insiemi a destra equivalgono al risultato della stringa di ricerca larynx OR nose NOT oral, mentre gli insiemi a sinistra per la stringa larynx OR (nose NOT oral); in entrambi le porzioni in grigio rappresentano il risultato della ricerca.

escludendo, fra queste, quelle che contengono oral) oppure larynx OR (nose NOT oral) (il motore restituirà tutte le citazioni che contengono larynx, compreso quelle che contengono anche oral, più quelle che hanno nose, escludendo fra queste, però, quelle che contengono oral) (vedere fig. 4).

PERFEZIONARE LA STRATEGIA DI BASE

Un ulteriore metodo per aumentare la specificità della propria ricerca è quello di utilizzare delle limitazioni di campo; ossia limitare la ricerca ad alcuni campi dei vari record (ricordiamo che i campi in Pubmed sono oltre 40 e contengono il nome dell'autore, il titolo della rivista, la data di pubblicazione, etc.).

Queste limitazioni possono essere impostate utilizzando i tag (etichette) dei campi fra parentesi quadre dopo la parola che si vuole limitare direttamente nel search box; ad esempio *larynx [TI]AND cancer [TI]* cercherà le due parole in combinazione solo nel titolo delle citazioni, attenzione poiché se scrivo *larynx AND cancer [TI]* non è la stessa cosa in quanto così il motore di ricerche mi restituirà tutte le citazioni in cui larynx compare in uno qualsiasi dei campi della citazione, mentre cancer solamente nel titolo.

Per rendere più semplice l'utilizzo delle limitazioni in Pubmed si può accedere alle stesse, in maniera molto intuitiva, dalla voce **limits** posta sopra il box di ricerca all'interno della pagina principale.

Alcuni suggerimenti per i limits:

- se si effettua una ricerca su patologie rare è consigliabile utilizzare l'intero database;
- se si vuole effettuare una revisione sistematica si può pensare di limitare la ricerca agli ultimi 5-6 anni;
- la limitazione di linguaggio ha poca importanza dal momento che la maggior parte delle citazioni presenti nel MEDLINE (come negli altri database) è in lingua inglese;
- al medico nella pratica clinica non servono gli articoli che trattano di procedimenti effettuati su animali per cui è consigliabile spuntare sempre la voce human nei limiti.

OPERAZIONI DI RICERCA AVANZATE

Fra le operazioni di ricerca avanzate, quelle che crediamo più importanti sono:

1. *Caratteri jolly*. Sfruttano i principi della word indexing sostituendo la parte finale di una parola. Vengono indicati in maniera differente a secondo del sistema utilizzato: nel PubMed il carattere jolly è l'asterisco: ad esempio, anem* ritroverà anemization, anemic, anemia, etc.⁶
2. *Subheadings*⁶. Utilizzandoli si può aumentare la specificità della ricerca. Tuttavia, poiché la loro assegnazione ha una riproducibilità molto bassa¹, bisogna usarli con attenzione dato che si corre il rischio di veder diminuire in maniera rilevante la sensibilità (possono sfuggire articoli interessanti).

FILTRARE I RISULTATI DELLA RICERCA

Nonostante tutti i nostri accorgimenti, per le inevitabili falle del sistema, i risultati delle nostre ricerche non saranno mai perfetti.

Questo comporta che, dopo aver effettuato una buona scrematura computerizzata, cioè dopo aver impostato bene la strategia di ricerca, i risultati ottenuti debbano essere rivisti "manualmente" attraverso la consultazione del titolo della citazione e, se disponibile, dell'abstract.

Naturalmente la rilevanza dei lavori è valutata in maniera soggettiva e variabile da caso a caso considerando anche gli obiettivi della ricerca.

Come regole generali pensiamo di poter suggerire di:

- Valutare sempre la rivista su cui l'articolo è pubblicato (riviste ad alto impact factor "dovrebbero" garantire lavori meglio condotti);
- Valutare il tipo di lavoro scientifico (naturalmente, ad esempio, un trial clinico randomizzato o meglio una revisione sistematica avranno più valore di uno studio retrospettivo);
- Valutare la data di pubblicazione (in medicina non ha molto senso

documentarsi su articoli troppo datati, poiché le conoscenze sono in continua evoluzione e, soprattutto in alcuni campi, bastano pochi anni per renderle obsolete).

Conclusioni

In conclusione possiamo affermare che nonostante tutti i limiti attuali dei database biomedici, questi abbiano portato notevoli benefici; in particolar modo hanno dato la possibilità a tutti di accedere ad una enorme quantità di dati cui attingere per i più svariati scopi (aggiornamento professionale, risoluzione di casi clinici, etc.).

Il rovescio della medaglia è che questa notevole quantità di dati deve essere adeguatamente filtrata per non correre il rischio di ottenere cattive informazioni che poi si risolvono in una scarsa qualità di cura per il paziente.

Bibliografia

1. Funk ME, Reid Ca. Indexing consistency in MEDLINE. Bull Med Libr Assoc 1983; 71: 176-83.
2. Gardner M. Information retrieval for patient care. BMJ 1997; 314: 950-3.
3. Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions. [Editorial] ACP J Club 1995; Nov/Dec: A-12. Ann Intern Med: 123, suppl 3.
4. Haynes RB, Wilczynski NL, McKibbon KA, Walker CJ, Sinclair JC. Developing optimal search strategies for detecting clinically sound studies in MEDLINE. J Am Med Inform Assoc 1994; 1: 447-58.
5. Cartabellotta A. Evidence-based Medicine 2. La ricerca dell'informazione biomedica clinicamente rilevante. Rec Prog Med 1998;89:265-74. (in http://www.gimbe.org/library/it/serie_articoli.asp)
6. Guida online sul sito <http://www.ncbi.nlm.nih.gov/pubmed/> in help ed advanced search
7. Home page sistema di indicizzazione MeSH <http://www.nlm.nih.gov/mesh/meshhome.html>
8. Cartabellotta A, per il Gruppo Italiano per la Medicina Basata sulle Evidenze — GIMBE®. Evidence-based Medicine 1. Il trasferimento dei risultati della ricerca alla pratica clinica. Rec Prog Med 1998; 89: 140-50.
9. Shaughnessy AF, Slawson DC, Bennet JH. Becoming an information master: a guidebook in the medical information jungle. J Fam Pract 1994; 39: 489-99.

**RICERCA ED ANALISI DELLE FONTI TRAMITE
BANCHE (DATI) ELETTRONICHE**

C. Vidale, S. Palma

I clinici si trovano spesso a dover eseguire ricerche bibliografiche, ma non sempre i risultati corrispondono alle attese.

Questo capitolo considera gli aspetti pratici della ricerca bibliografica biomedica su banche dati elettroniche disponibili in internet. In particolar modo si affrontano i metodi di ricerca bibliografica in PubMed. Il capitolo rappresenta un contributo sia alla necessità di aggiornamento richiesto dalla professione medica, sia al bisogno di rispondere ai quesiti clinici che si presentano quotidianamente.

Eseguire ricerche bibliografiche che producono risultati pertinenti e validi richiede un approccio metodologico corretto.

Inoltre, poiché l'economia del lavoro non sempre consente di dilungarsi in una dettagliata analisi dei risultati, a volte si leggono esclusivamente i riassunti prodotti dalla ricerca. Impostando invece una ricerca bibliografica corretta dal punto di vista metodologico già all'origine, il numero dei risultati potrà essere sufficientemente limitato, tanto da permettere l'accesso al testo pieno dei documenti effettivamente pertinenti ed il più possibile specifici. Tutto ciò senza dispendio inutile di tempo.

Per ogni medico otorinolaringoiatra è utile prendere contatti con la biblioteca della propria Azienda Sanitaria, al fine di conoscere quali banche dati, riviste elettroniche e cartacee e libri sono presenti a catalogo e, pertanto, conoscere la consistenza del patrimonio accessibile.

Si tenga presente che non tutte le riviste di otorinolaringoiatria sono disponibili in formato elettronico e che può essere necessario recarsi fisicamente in biblioteca per consultare riviste non disponibili nella rete.

La panoramica delle fonti prese in considerazione si basa principalmente su risorse elettroniche gratuite o che possono essere interrogate liberamente, ma alle quali l'editore inibisce l'accesso ai documenti a testo completo.

Da una consultazione¹ nel portale dell'Istituto Superiore di Sanità dedicato al progetto PIRAMIDE, (<http://www.iss.it/pira/>) l'emeroteca virtuale dei periodici indicizzati da PubMed e presenti nelle biblioteche italiane, emerge che PubMed censisce 119 riviste (attive e cessate) in ambito otorinolaringoiatrico e che una buona parte delle stesse (aggiornamento

¹ Consultazione del 3 giugno 2010

2008) non è disponibile in formato elettronico. Non tutte le informazioni, dunque, sono presenti nella rete, ma ciò non deve portare alla conclusione errata che l'informazione cercata non esiste. In questi casi si deve consultare il catalogo del materiale cartaceo che ogni biblioteca mette a disposizione dell'utente.

Inoltre bisogna considerare che nel web si possono trovare moltissime informazioni, ma non sempre le fonti sono certificate. Per ovviare a questa difficoltà, si propongono alcuni accorgimenti che permettono, in linea generale, di non cadere in tranelli informativi. Essi sono:

Scegliere solo risorse certificate. Il bollino dell'Health on the Net (HON-code <http://www.hon.ch/HONcode/Italian/>) può essere un valido aiuto;

- Considerare solo siti/risorse/banche dati i cui contenuti sono dichiarati e di cui si conoscono gli autori/editori, gli eventuali sponsor, le date d'aggiornamento ecc.;
- Interrogare i motori di ricerca Google, Yahoo ecc., valutando sempre la provenienza e l'affidabilità dei dati e dei siti emersi dalla ricerca;
- Utilizzare i motori di ricerca scientifici come Google Scholar (<http://scholar.google.it/>) e Scirus (<http://scirus.com/>);
- Basare la propria ricerca su siti/risorse/banche dati Evidence Based Medicine, come ad esempio The Cochrane Library (<http://www.thecochranelibrary.com>), TRIP (Translating Research Into Practice) database (<http://www.tripdatabase.com/>), NHS Evidence <http://www.evidence.nhs.uk/default.aspx> ed il suo derivato DUETs (Database of Uncertainties about the Effects of Treatments) <http://www.library.nhs.uk/DUETs/>.
- Appoggiarsi a PubMed per quanto riguarda la ricerca basata su fonti primarie.

Le risorse citate consentono l'accesso gratuito ai dati bibliografici e sono, pertanto, disponibili a tutti gli utenti.

In commercio esistono numerose risorse di supporto alle decisioni cliniche che esulano da questo breve contributo.

L'oggetto delle prossime pagine è, appunto, l'orientamento nella ricerca in PubMed. Le esercitazioni proposte utilizzano la piattaforma originale della National Library of Medicine e si basano su una solida esperienza nella ricerca bibliografica e consentono di ottenere risultati mirati.

PubMed è un'interfaccia web sviluppata dal National Center for Biotechnology Information (NCBI), settore della National Library of Medicine (NLM), parte del National Institute of Health (NIH) degli Stati Uniti d'America.

PubMed (significa Public MEDLINE, MEDLINE a sua volta è l'acronimo di MEDLARS—Medical Literature Analysis and Retrieval System—on line)

è accessibile gratuitamente dalla collettività degli utenti dal 26 giugno 1997. È un database di citazioni bibliografiche usato internazionalmente e prevede l'accesso alla letteratura biomedica mondiale. La lingua utilizzata è l'inglese.

La National Library of Medicine, attraverso il Literature Selection Technical Review Committee–LSTRC (<http://www.nlm.nih.gov/pubs/factsheets/jsel.html>), provvede alla selezione dei titoli dei periodici da inserire nella banca dati in base a criteri standardizzati e trasparenti.

Si tenga presente che, trattandosi di citazioni bibliografiche tratte da riviste primarie, non è prevista, da parte della National Library of Medicine, alcun'analisi critica del contenuto degli articoli.

Al momento della stesura di questo contributo sono indicizzati 5455 periodici² per i quali è stato fatto lo spoglio di oltre 18.000.000 di articoli.

La ricerca in PubMed

Lo scopo di questa sezione è quello di semplificare in modo schematico, per quanto possibile, i metodi per condurre una corretta ricerca bibliografica evitando spiegazioni su modalità di utilizzo più complesse. Così facendo il medico otorinolaringoiatra si troverà di fronte ad un percorso orientato, ripetibile per la maggior parte delle ricerche bibliografiche biomediche.

PubMed è un sistema duttile che permette di conseguire risultati utilizzando percorsi diversi.

Le spiegazioni, sempre molto sintetiche, sono supportate da illustrazioni, in modo da dare al lettore la possibilità di ripetere i percorsi proposti e cimentarsi con le proprie abilità.

PubMed prevede più modi di fare ricerca, ognuno dei quali racchiude caratteristiche peculiari. L'approccio è diverso a seconda del tipo di ricerca individuato, come diversi sono i risultati secondo la *query*³ proposta (vedi anche cap. 9). Spesso, se si desidera ottenere dati piuttosto esaustivi, la somma di più ricerche risulta essere la più completa.

In questo capitolo si prende in esame, a titolo esclusivamente esemplificativo, il seguente quesito: Qual è la terapia nelle neoplasie secon-

² Per individuare il numero di periodici indicizzati alla data corrente: accedere al sito. <http://www.ncbi.nlm.nih.gov/journals> e digitare nel riquadro di ricerca "currentlyindexed[All Fields]".

³ L'insieme di termini utilizzati per definire i concetti della ricerca.

darie faringee? Detto quesito è riproposto con i seguenti percorsi di ricerca:

- “PubMed basic search”;
- “MeSH Database”;
- “Search Builder”;
- “Clinical Queries”.
- Il primo passo per impostare una strategia di ricerca (carta e penna alla mano!) è la pianificazione del bisogno informativo⁴: si tratta quindi di definire la strategia di ricerca per dare risposta al quesito. Per fare ciò il quesito deve essere scomposto in tutte le sue parti e, per quanto riguarda il tema oggetto dello studio, esattamente:
 - Neoplasie secondarie faringee;
 - Terapia di qualsiasi tipo (ossia: farmacologica, chirurgica ecc.) legata alle neoplasie secondarie faringee.

Bisognerà inoltre:

- Tradurre i termini/concetti in lingua inglese;
- Individuare eventuali sinonimi;
- Stabilire la relazione che i diversi termini/concetti hanno fra loro tenendo presente le caratteristiche per l’uso degli operatori booleani AND, OR e qualche rara volta NOT (vedi cap. 9).

PubMed basic search

È il tipo di ricerca più diffuso, ma anche quello generalmente meno efficace, se non si conoscono bene i meccanismi di ricerca bibliografica. È basato su un sistema di conversione di termini (*Automatic Term Mapping*) che “interpreta” le parole inserite nel riquadro di ricerca a seguito di algoritmi predefiniti. In questo caso il medico assume un ruolo passivo. Si sceglie di utilizzare i termini inglesi: *Secondary Pharyngeal Neoplasm Therapy* che vanno inseriti nel riquadro di ricerca della pagina di apertura di PubMed (Fig. 1).

Il sistema converte la query⁵ in:

(“secondary”[Subheading] OR “secondary”[All Fields]) AND (“pharyngeal neoplasms”[MeSH Terms] OR (“pharyngeal”[All Fields] AND “neoplasms”[All Fields]) OR “pharyngeal neoplasms”[All Fields] OR

⁴ Per scaricare template utili per la ricerca: *La ricerca medica nel web: modelli e stili*. 2008 <http://eprints.rclis.org/14731/>.

⁵ Per verificare la conversione della query da parte dell’*Automatic Term Mapping* bisogna leggere quanto scritto in “Search details” che si trova nella videata in basso a destra della pagina.

Google™ Questa pagina è in inglese. Tradurre la pagina con Google Toolbar? Ulteriori informazioni. Non in inglese? Aiutaci Traduci Traduci sempre Inglese

PubMed
 U.S. National Library of Medicine
 National Institutes of Health

Search: PubMed
 Secondary Pharyngeal Neoplasm Therapy
 Limits Advanced search Help
 Search Clear

PubMed
 PubMed comprises more than 19 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals Database
- Clinical Trials
- E-Utilities
- LinkOut

Getting Started

- NCBI Help Manual
- NCBI Handbook
- Training & Tutorials

Resources

- Literature
- DNA & RNA
- Proteins

Popular

- Pubmed
- Nucleotide
- BLAST

Featured

- GenBank
- Reference Sequences
- Map Viewer

NCBI Information

- About NCBI
- Research at NCBI
- NCBI Newsletter

You are here: NCBI > Literature > PubMed

Write to the Help Desk

Fig. 1

("pharyngeal"[All Fields] AND "neoplasm"[All Fields]) OR "pharyngeal neoplasm"[All Fields] AND ("therapy"[Subheading] OR "therapy"[All Fields] OR "therapeutics"[MeSH Terms] OR "therapeutics"[All Fields]).

Le citazioni bibliografiche prodotte sono 1044⁶.

Mesh Database

Il "MeSH Database" è il sistema portante della ricerca bibliografica in PubMed. Rappresenta la ricerca concettuale tendente ad ottenere citazioni bibliografiche (che fanno riferimento ad articoli di riviste) riguardanti l'argomento desiderato. "MeSH" significa *Medical Subject Heading*, ossia "descrittori di carattere medico". Il "MeSH Database" è un vocabolario controllato (detto anche *thesaurus*) di termini medici. Per ogni concetto è presente un termine/concetto (spesso rappresentato da più parole) nel quale si riconoscono tutti i sinonimi. In pratica, esemplificando, il sistema riconduce termini come *cancer, tumour, tumor, tumors, neoplasm, neoplasms* al concetto unico di *neoplasms*.

Gli articoli presenti in PubMed sono per la gran maggioranza indicizzati, ossia sono provvisti di chiavi d'accesso utili per il loro reperimento. Esistono due tipi di indicizzazione: uno è derivato dai termini presenti nel testo pubblicato da PubMed, l'altro è attribuito dai bibliotecari documentalisti biomedici della National Library of Medicine. Per assegnare i "MeSH" agli articoli presenti in PubMed, si utilizza il vocabolario controllato, o *thesaurus*. Nel caso in cui un articolo non sia stato ancora indicizzato (in linea di massima trattasi di informazioni non più vecchie di uno, due mesi o prodotte direttamente dall'editore) non è rintracciabile con la ricerca "MeSH Database".

Fatta questa doverosa premessa, si riparte dal quesito citato e già scomposto nelle sue singole parti:

- Neoplasie secondarie faringee
- Terapia di qualsiasi tipo (ossia: farmacologica, chirurgica ecc.) delle neoplasie secondarie faringee.

Si apre la pagina del "MeSH Database"⁷ e si digita nel *box* di ricerca: *Pharyngeal neoplasms*. Come premesso è indifferente digitare *Pharyngeal cancer, Pharyngeal tumor, Pharyngeal neoplasm*⁸. Il vocabolario controllato riconduce il termine di ricerca a *Pharyngeal Neoplasms* (Fig. 2).

⁶ Dati aggiornati al 17 giugno 2010.

⁷ Si trova nella terza colonna della pagina di apertura di PubMed sotto "More Resources".

⁸ Non cambia se i termini sono inseriti al plurale.

Appare una schermata con evidenziato in grassetto *Pharyngeal Neoplasms*, la definizione del concetto, un numero di Subheadings⁹, la possibilità di restringere la ricerca (“Restrict Search to Major Topic headings only”) la possibilità di non esplodere la ricerca (“Do Not Explode this term”)¹⁰, gli “Entry Terms”¹¹ e la struttura gerarchica strutturata ad albero composta di quattro rami, più foglie e foglioline¹² (Fig. 3).

Tutti i termini/concetti dell’albero sono linkati con la pagina che li riguarda direttamente.

Il quesito preso in considerazione richiede di evidenziare l’aspetto terapeutico della neoplasia faringea secondaria.

PubMed attribuisce al “subheading” *secondary* il compito di evidenziare le neoplasie di origine secondaria e a quello di *therapy* quello di evidenziare tutti i tipi di terapia¹³.

Si sceglierà quindi di spuntare i “Subheadings” *therapy* e *secondary* e di avviare la ricerca¹⁴ (Fig. 4)

La sintassi della *query* è pertanto:

(“Pharyngeal Neoplasms/secondary”[Mesh] AND “Pharyngeal Neoplasms/therapy”[Mesh]).

Le citazioni bibliografiche prodotte sono 99.

Volendo restringere il numero dei risultati verso una ricerca più specifica, si può spuntare il quadratino che precede “Restrict Search to Major Topics heading only”¹⁵ ed eventualmente escludere tutti i termini/concetto subordinati e sottostanti a *Pharyngeal Neoplasms*, spuntando il quadratino che precede l’indicazione di “Do Not Explode this term”¹⁶.

⁹ I “Subheading” sono qualificatori di “MeSH” e servono ad evidenziare un particolare aspetto del descrittore. Il loro numero è variabile a seconda del concetto al quale si riferiscono.

¹⁰ PubMed prevede l’esplosione automatica di tutti i concetti dipendenti da quello scelto, ossia estende la ricerca a tutti i concetti gerarchicamente subordinati al quello scelto. In questo caso l’esplosione comprende: *Hypopharyngeal Neoplasms, Nasopharyngeal Neoplasms, Oropharyngeal Neoplasms e Tonsillar Neoplasms*.

¹¹ Sono i termini che per PubMed corrispondono a quello digitato.

¹² Nell’immagine sono evidenziati due rami della struttura ad albero.

¹³ *Therapy* comprende: *diet therapy, drug therapy, nursing, prevention and control, radiotherapy, rehabilitation, surgery e transplantation*.

¹⁴ I passaggi dopo aver spuntato il quadratino antistante *therapy* e *secondary* sono i seguenti: Nella tendina di “Send to” scendere col cursore fino a “Search Box with AND”. Dare l’invio. In automatico appariranno i termini di ricerca trasferito nel box di attesa concatenati con l’operatore logico OR. Cancellare l’operatore logico OR e sostituirlo con l’operatore logico AND. Cliccare su “Search PubMed”. Vengono evidenziati i lavori corrispondenti alla strategia di ricerca.

¹⁵ In questo caso la sintassi è: *(“Pharyngeal Neoplasms/secondary”[MAJR] AND “Pharyngeal Neoplasms/therapy”[MAJR])* e produce 12 citazioni bibliografiche.

¹⁶ In questo caso la sintassi è: *“Pharyngeal Neoplasms/secondary”[Mesh:noexp] AND “Pharyngeal Neoplasms/therapy”[Mesh:noexp]* che non produce citazioni bibliografiche.

Google™ Questa pagina è in inglese. Tradurre la pagina con Google Toolbar? Ulteriori informazioni Non in Inglese? Aiutaci a migliorare

Traduci Traduci sempre Inglese

All MeSH Categories
Diseases Category
Neoplasms
 Neoplasms by Site
 Head and Neck Neoplasms
 Otorhinolaryngologic Neoplasms
 Pharyngeal Neoplasms
 Hypopharyngeal Neoplasms
 Nasopharyngeal Neoplasms
 Oropharyngeal Neoplasms
 Tonsillar Neoplasms

All MeSH Categories
Diseases Category
Stomatognathic Diseases
 Pharyngeal Diseases
 Pharyngeal Neoplasms
 Hypopharyngeal Neoplasms
 Nasopharyngeal Neoplasms
 Oropharyngeal Neoplasms
 Tonsillar Neoplasms

All MeSH Categories
Diseases Category
Otorhinolaryngologic Diseases
 Otorhinolaryngologic Neoplasms
 Pharyngeal Neoplasms
 Hypopharyngeal Neoplasms
 Nasopharyngeal Neoplasms
 Oropharyngeal Neoplasms
 Tonsillar Neoplasms

All MeSH Categories
Diseases Category
Otorhinolaryngologic Diseases
 Pharyngeal Diseases

Fig. 3

A service of the National Library of Medicine and the National Institutes of Health

My NCBI [Sign In] [Register]

MeSH

All Databases: Search [MeSH] for pharyngeal cancer

Published Nucleotide Protein Genome Structure OMIM PMC Journals Books

Limits Preview/Index History Clipboard Details

Search PubMed Clear

Suggestions: Pharyngeal cancer, Pharyngeal cancers, Cancer, pharyngeal, Laryngeal cancer, Laryngeal cancers, Cancers, pharyngeal, Oropharyngeal cancer, Canc laryngeal, Pharyngeal cleft, Pharyngeal clefts, More...

Display [Full] Show 20 Send to

All: 1

- If making selections (e.g., Subheadings, etc.), use the Send to Search Box feature to see PubMed records with those specifications.
- Select PubMed under the Links menu to retrieve all records for the MeSH Term.
- Select NLM MeSH Browser under the Links menu for additional information.

1: Pharyngeal Neoplasms
Tumors or cancer of the PHARYNX.

Subheadings: This list includes those paired at least once with this heading in MEDLINE and may not reflect current rules for allowable combinations.

- analysis
- blood
- blood supply
- chemically induced
- chemistry
- classification
- complications
- congenital
- diet therapy
- drug therapy
- economics
- embryology
- enzymology
- epidemiology
- ethnology
- etiology
- genetics
- history
- immunology
- metabolism
- microbiology
- mortality
- nursing
- parasitology
- pathology
- physiology
- pathophysiology
- prevention and control
- psychology
- radiography
- radionuclide imaging
- radiotherapy
- rehabilitation
- secondary
- secretion
- surgery
- therapy
- ultrasonography
- ultrastructure
- urine
- veterinary
- virology

Restrict Search to Major Topic headings only.

Do Not Explode this term (i.e., do not include MeSH terms found below this term in the MeSH tree).

Entry Terms:

Related Resources: Order Documents, NLM Mobile, NLM Catalog, NLM Gateway, TOXNET, Consumer Health, Clinical Alerts, ClinicalTrials.gov, PubMed Central

Fig. 4

Sono possibili anche abbinamenti fra “Restrict Search to Major Topics heading only” e “Do Not Explode this term”

Search Builder

È una modalità di ricerca basata su parole specifiche, corredate da etichette che definiscono il campo di appartenenza.

“Search Builder” si trova nella pagina dell’ “Advanced Search” di PubMed ed è uno strumento per impostare ricerche bibliografiche costruite sui termini citati. Con tale opzione è possibile evidenziare immediatamente il numero delle ricorrenze del termine presente nei campi del database. Si basa sul principio di appartenenza di una determinata parola ad un campo specifico.

Si inizia dal quesito *Secondary Pharyngeal Neoplasm Therapy*. Si possono fare diverse scelte e per ogni scelta valutarne l’opportunità. I campi più utilizzati sono quelli definiti come “Author”, “Title”, “Abstract”, “MeSH” e “All Fields”. Nel nostro esempio scegliamo di cercare *Pharyngeal Neoplasms* in tutti i campi e aggiungere i termini *therapy* e *secondary* nel titolo e nell’abstract (Fig. 5).

Alla fine dell’inserimento dei termini¹⁷, la sintassi della query è: (“*pharyngeal neoplasms*”[All Fields]) AND *therapy*[Title/Abstract]) AND “*secondary*”[Title/Abstract].

Le citazioni bibliografiche prodotte sono 21.

Questa stringa rappresenta un’associazione di parole e di termini/concetti¹⁸ presenti in campi differenti del database. In questa ricerca i termini *therapy* e *secondary* non hanno legami con *Pharyngeal Neoplasms*, ma sono parole presenti nel campo “Title/Abstract”.

Si ricorda a tal proposito la strategia di ricerca con il “MeSH Database”, dove i “Subheadings” *secondary* e *therapy* rappresentano aspetti specifici della neoplasia faringea.

¹⁷ I passaggi sono i seguenti: digitare i termini nel riquadro di ricerca “Search Builder” e selezionare il campo di ricerca. Cliccare su Show index e successivamente evidenziare il termine. Cliccare su Add to Search Box per inviare il termine nella Search Box. Proseguire per gli altri termini nello stesso modo verificando sempre di aver inserito l’operatore logico corretto.

¹⁸ A tal proposito si ricorda che *Pharyngeal Neoplasms* è anche un termine “MeSH”. Pertanto, effettuando la ricerca in tutti i campi si include anche il campo del “MeSH Database”.

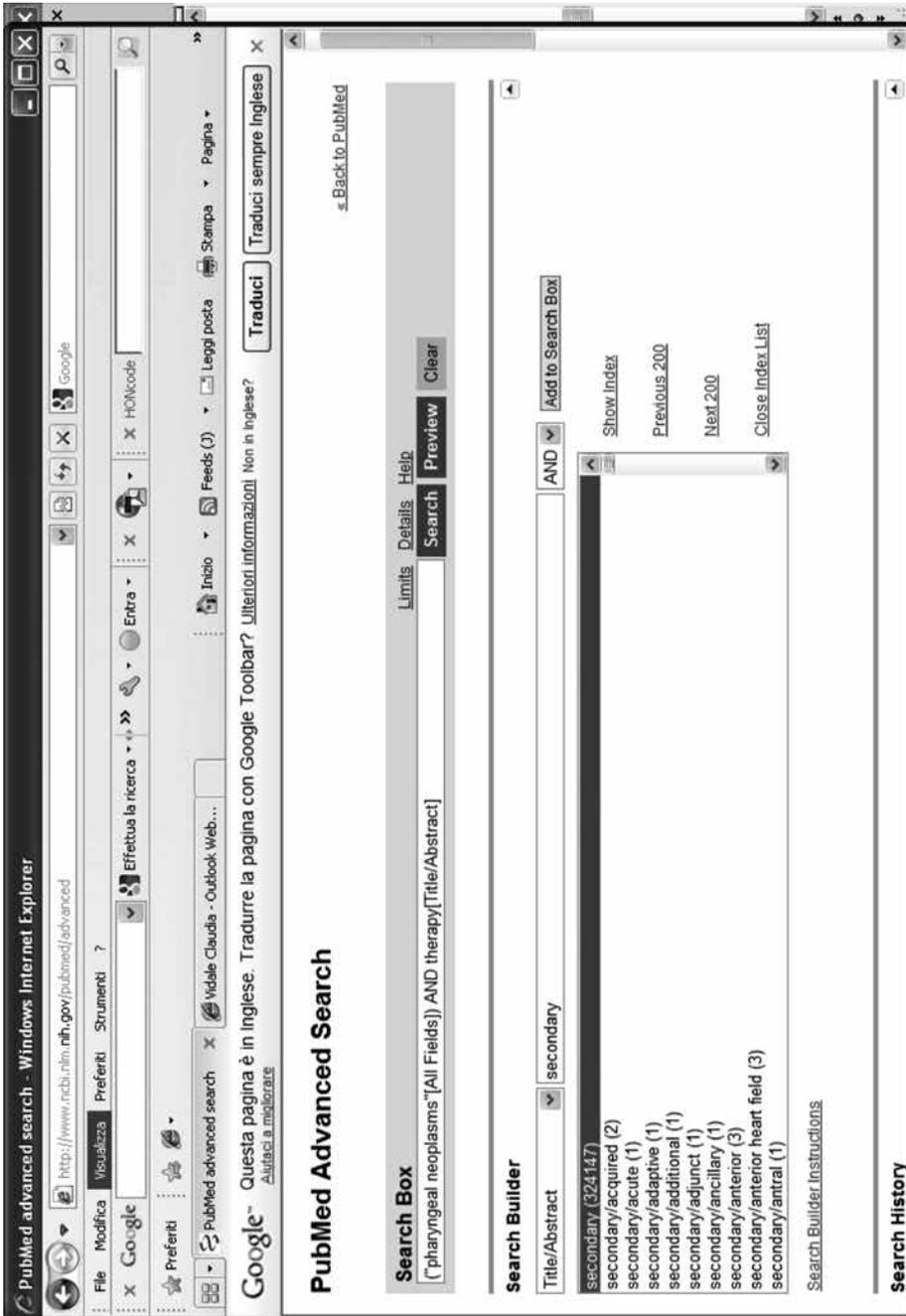


Fig. 5

Clinical Queries

È un tipo di ricerca basata sui principi dell'Evidence Based Medicine¹⁹. Come per "PubMed basic search" il medico si lascia condurre dal sistema per ottenere i risultati bibliografici.

Si apre la pagina delle "Clinical Queries"²⁰ dove appaiono²¹:

- "Clinical Study Categories";
- "Systematic Reviews";
- "Medical Genetics".

Con la medesima *query* si ottengono i risultati per tutte le tre "Clinical Queries".

Si digita nella casella di ricerca i termini *Secondary Pharyngeal Neoplasm* e si seleziona la "Category" *therapy* (termine evidenziato di *default* nella tendina).

Si può scegliere di indirizzare la ricerca a risultati più limitati e specifici, selezionando il termine *narrow* nella tendina "Scope", oppure a risultati più ampi e sensibili, lasciando nella tendina "Scope" il termine *broad* proposto di *default*.

Di seguito si evidenzia la pagina dei risultati²² (Fig. 6).

Di seguito si evidenzia il prospetto riepilogativo delle ricerche effettuate, con esclusione delle "Clinical Queries" già evidenziato in precedenza, che appare in "Search History"²³ (Fig. 7).

I risultati delle ricerche che appaiono nella sezione "Search History" possono essere concatenati fra loro con gli operatori logici.

Un abile ricercatore è in grado di costruire, direttamente da "PubMed basic search", tutti i percorsi proposti attribuendo le diverse etichette (o *tag*²⁴). Si consiglia detta procedura esclusivamente a chi ha un'approfondita conoscenza del sistema.

I risultati confermano, o meno, la validità della ricerca bibliografica, tenendo presente che qualsiasi scelta esclude un certo numero di lavori, maggiori se il bisogno è quello di ottenere dati specifici.

¹⁹ Si basa su un convertitori automatici e filtri studiati per agevolare la difficoltà nella costruzione del quesito clinico.

²⁰ Il link si trova nella seconda colonna della pagina di apertura di PubMed sotto "PubMed Tools".

²¹ La pagina di PubMed Clinical Queries è stata ridisegnata a luglio 2010.

²² Dati aggiornati al 16 luglio 2010.

²³ "Search History" si trova nell' "Advanced Search" a sua volta raggiungibile dalla pagina principale di PubMed.

²⁴ Per *tag* o etichetta si intende l'abbreviazione presente fra le parentesi quadre ossia [MeSH], [All Fields], [sb] e può riferirsi all'indicazione di campi, limiti, sottoinsiemi (*subset*) ecc.

Google® Questa pagina è in Inglese. Tradurre la pagina con Google Toolbar? Ulteriori informazioni Non in Inglese? Aiutaci a migliorare

NCBI Resources How To Traduci My NCBI Sign In

PubMed Clinical Queries

Search Secondary Pharyngeal Neoplasms Search Clear

Results of searches on this page are limited to specific clinical research areas. For comprehensive searches, use PubMed directly.

Clinical Study Categories

Category: Therapy

Scope: Narrow

Systematic Reviews

Results: 5 of 14

Secondary tracheoesophageal puncture with in-office transnasal esophagoscopy. [Arch Otolaryngol Head Neck Surg. 2009]

Joint practice guidelines for radionuclide lymphoscintigraphy for sentinel node localization in oropharyngeal squamous c [Am Surg Oncol. 2009]

Larynx preservation clinical trial design: key issues and recommendations—a consensus panel summary. [Int J Radiat Oncol Biol Phys. 2009]

Larynx preservation clinical trial design: key issues and recommendations—a consensus panel summary. [Head Neck. 2009]

Role of chest CT in staging of oropharyngeal cancer: a systematic review. [Head Neck. 2009]

See all (14)

Filter citations for systematic reviews, meta-analyses, reviews of clinical trials, evidence-based medicine, consensus development conferences, and guidelines. See related sources.

Medical Genetics

Topic: All

Results: 5 of 96

[North African and Southeast Asian nasopharyngeal carcinomas: between the resemblance and the dissemblance] [Bull Cancer. 2010]

Functional polymorphism of cytotoxic T-lymphocyte antigen 4 and nasopharyngeal carcinoma susceptibility in a Chinese population. [Int J Immunogenet. 2010]

Lysyl oxidase expression is an independent marker of prognosis and a predictor of lymph node metastasis in oral and oropharyngeal squamous [Int J Cancer. 2010]

Differential baseline and response profile to FN-gamma gene transduction of IL-6/IL-6 receptor-alpha secretion discriminate primary tum [BMC Cancer. 2009]

Altered desmoplakin expression at transcriptional and protein levels provides prognostic information in human oropharyngeal cancer. [Hum Pathol. 2009]

See all (96)

Filter citations to topics in medical genetics.

Fig. 6

Google™ Questa pagina è in inglese. Tradurre la pagina con Google Toolbar? Ulteriori informazioni Non in inglese? Aiutaci Traduci sempre Inglese Traduci

NCBI Resources How To My NCBI Sign In

PubMed Advanced Search

s. Back to PubMed

Search Box Limits Details Help Search Preview Clear

Search Builder All Fields AND Add to Search Box Show Index

Search Builder Instructions

Search History Most Recent Queries

Search	Time	Result
#8 Search (Secondary Pharyngeal Neoplasms Therapy) AND systematic(sb)	08:56:41	10
#7 Search (Secondary Pharyngeal Neoplasms) AND (Therapy/narrow(filter))	08:55:51	24
#6 Search ("pharyngeal neoplasms"[All Fields]) AND ("therapy"[Title/Abstract]) AND "secondary"[Title/Abstract]	08:45:32	21
#5 Search ("Pharyngeal Neoplasms/secondary"[MAJR:HoExp] AND "Pharyngeal Neoplasms/therapy"[MAJR:HoExp])	08:44:50	0
#4 Search ("Pharyngeal Neoplasms/secondary"[Mesh:HoExp] AND "Pharyngeal Neoplasms/therapy"[Mesh:HoExp])	08:44:15	4
#3 Search ("Pharyngeal Neoplasms/secondary"[MAJR] AND "Pharyngeal Neoplasms/therapy"[MAJR])	08:43:39	12
#2 Search ("Pharyngeal Neoplasms/secondary"[Mesh] AND "Pharyngeal Neoplasms/therapy"[Mesh])	08:42:55	99
#1 Search Secondary Pharyngeal Neoplasm Therapy	08:39:35	1044

Less History Clear History Search History Instructions

More Resources
 MeSH Database
 Journals Database
 Simple Citation Matcher

Fig. 7

In queste poche pagine di orientamento alla ricerca in PubMed, si è scelto di non utilizzare i "Limits" (temporali, di pubblicazione, di genere, di età ecc.) messi a disposizione dal programma. Circostringere la propria ricerca secondo parametri principalmente temporali o semplificando troppo il tema dell'indagine spesso porta a risultati non pertinenti e troppo sensibili.

Ottimi risultati di qualità si possono ottenere sfruttando anche l'opportunità di PubMed per individuare le citazioni delle revisioni sistematiche della Cochrane Library <http://www.thecochranelibrary.com/>²⁵.

In questo caso l'esempio sviluppato in precedenza rappresenta una situazione troppo di nicchia per essere presente nelle revisioni sistematiche Cochrane. Come noto la metodologia richiesta per produrre una revisione sistematica Cochrane è molto complessa e prevede anche un'attenta valutazione della letteratura pubblicata e non pubblicata²⁶. Il numero di revisioni sistematiche e dei protocolli²⁷ presenti nel database sono esigui se confrontati con i grandi numeri di PubMed (Fig. 8).

Per questa ragione la *query* utilizzata in questo caso prende in considerazione esclusivamente *Pharyngeal Neoplasms*, senza tener conto dell'aspetto della terapia e della neoplasia di carattere secondario.

Si ripropone, in questo caso²⁸, la modalità "Search Builder" costruendo una stringa di ricerca dove sono contenuti i termini *cochrane database of systematic reviews* e alternativamente sono presenti *Pharyngeal Neoplasms* nel campo del descrittore "MeSH" oppure nel titolo e/o abstract²⁹.

La sintassi della *query* effettuata è: "*cochrane database of systematic reviews online*"[All Fields] AND ("*pharyngeal neoplasms*"[MeSH Terms] OR "*pharyngeal neoplasms*"[Title/Abstract]).

Le citazioni bibliografiche prodotte sono 2 e vengono di seguito riportate (Fig. 9).

Il medico può cliccare sul titolo della citazione per ottenere l'abstract e collegarsi al sito della Cochrane Library o altra piattaforma in abbonamento (es. OvidSp, EBSCOHost) per ottenere l'articolo in *full text*.

²⁵ Si rinvia al sito originale per l'approfondimento. Si ricorda che è necessario sottoscrivere l'abbonamento per leggere il testo pieno delle revisioni sistematiche.

²⁶ Questo tipo di letteratura che non rientra nei canali consueti dell'editoria si definisce Letteratura grigia

²⁷ Alla pagina <http://www.thecochranelibrary.com/view/0/AboutTheCochraneLibrary.html> della Cochrane Library alla data del 28 maggio 2010 si dichiarano 6162 articoli di cui 4247 revisioni e 1915 protocolli.

²⁸ È possibile effettuare la ricerca anche con altri percorsi.

²⁹ In questi casi è bene fare molta attenzione al concatenamento dei termini e all'inserimento delle parentesi. PubMed processa prima i termini collegati con l'operatore logico AND e successivamente quelli con l'operatore logico OR. Se non vengono inserite le parentesi i risultati saranno completamente diversi e non pertinenti.

The screenshot shows a web browser window titled "PubMed advanced search - Windows Internet Explorer". The address bar shows the URL: <http://www.ncbi.nlm.nih.gov/pubmed/advanced>. The browser's search bar contains the text "PubMed advanced search".

The main content area is titled "PubMed Advanced Search" and features a "Search Box" with the following query: `[\"cochrane database of systematic reviews online\"[All Fields]] AND \"pharyngeal neoplasr\"`. Below the search box are buttons for "Limits", "Details", "Help", "Search", "Preview", "Clear", and "Add to Search Box".

The "Search Builder" section shows the selected search terms: "Title/Abstract" and "pharyngeal neoplasm*".

The search results are displayed in a list format:

- pharyngeal neoplasm (5)
- pharyngeal neoplasms (24)
- pharyngeal nerve (27)
- pharyngeal nerve plexus (3)
- pharyngeal nerve ring (7)
- pharyngeal nerves (15)
- pharyngeal nervous system (16)
- pharyngeal neuralgia (7)
- pharyngeal neurectomy (4)
- pharyngeal neuron (4)

At the bottom of the results list, there are links for "Show Index", "Previous 200", "Next 200", and "Close Index List".

The browser's status bar at the bottom indicates "My NCBI Sign in".

Fig. 8

The screenshot shows a web browser window with the following content:

- Address Bar:** <http://www.ncbi.nlm.nih.gov/sites/pubmed>
- Search Bar:** Search: PubMed. Query: "cochrane database of systematic reviews online"[All Fields] AND ("pharyngeal neoplasms"[MeSH T - Windows Internet Explorer]
- Search Results:**
 - Results: 2**
 - Interventions for the treatment of oral and oropharyngeal cancers...surgical treatment.
 - 1. Oliver RJ, Clarkson JE, Conway DI, Glenny A, MacLuskey M, Pavitt S, Sloan P, CSROC Expert Panel, Worthington HV. Cochrane Database Syst Rev. 2007 Oct 17;(4):CD006205. Review. PMID: 17943894 [PubMed - indexed for MEDLINE] Related citations
 - Chemotherapy as an adjunct to radiotherapy in locally advanced nasopharyngeal carcinoma.
 - 2. Baujat B, Audry H, Bourhis J, Chan AT, Onat H, Chua DT, Kwong DL, Al-Sarraf M, Chi KH, Hareyama M, Leung SF, Thephamongkhon K, Pignon JP, MAC-NPC Collaborative Group. Cochrane Database Syst Rev. 2006 Oct 18;(4):CD004329. Review. PMID: 17054200 [PubMed - indexed for MEDLINE] Related citations
- Filter your results:**
 - All (2)
 - Review (2)
 - Free Full Text (0)
- Find related data:** Database: Select. Find items
- Search details:** "cochrane database of systematic reviews online"[All Fields] AND ("pharyngeal neoplasms"[MeSH Terms] OR "pharyngeal

Fig. 9

Tutte le ricerche possono essere salvate utilizzando l'archivio permanente "My NCBI"³⁰.

La ricerca perfetta probabilmente non esiste. È possibile, però, raggiungere ottimi risultati seguendo poche regole di base in modo metodico, combinando i dati e le citazioni ottenute attraverso più ricerche.

Per saperne di più

- PubMed Tutorials <http://www.nlm.nih.gov/bsd/disted/pubmed.html>.
- PubMed Quick Start Guide http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helppubmed&part=pubmedhelp#pubmedhelp.PubMed_Quick_Start.
- PubMed Training Manual http://www.nlm.nih.gov/pubs/web_based.html.
- PubMed Help <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helppubmed&part=pubmedhelp>.
- Ceccarini A, Falcone MA, Ferrari P. NLM Italian translation: evolution in progress. *Journal of the European Association for Health Information and Libraries*. 2008; 4(1):3-6. http://www.eahil.net/journal/journal_2008_vol4_n1.pdf.
- Corrao S. *Conoscere e usare PubMed. Guida al più noto sistema di ricerca bibliografica in campo sanitario*. Roma, Il Pensiero Scientifico, 2 ed. 2008.

³⁰ A tal proposito si invia alle informazioni presenti al collegamento "My NCBI" <http://www.ncbi.nlm.nih.gov/sites/myncbi/about/> evidenziato in alto a destra di ogni pagina di PubMed.

APPENDICE
LA TABELLA 2 X 2 UNIVERSALE
S. Palma

Proviamo ora a costruire una tabella 2 x 2 universale, ovvero utile sia per i test diagnostici che per la terapia o rischio di esposizione, utilizzando un foglio elettronico di Excel.

Da A1 a D4 tracciamo una griglia, le caselle della tab. 2 x 2 sono costituite da B2, C2, B3 e C3. In A2 indichiamo gli esposti (o trattati) o i soggetti positivi al test, in A3 i soggetti non esposti (o non trattati) o i soggetti negativi al test, in B1 e C1 indichiamo rispettivamente i soggetti affetti dalla malattia (M+) e i soggetti non ammalati (M-) (fig. 1).

In D2 scriviamo la seguente formula: =B2+C2 (N.B. il segno "=" indica l'inserimento di una formula); in D3 scriviamo la seguente formula: =B3+C3. In B4 indichiamo la somma della colonna B: =B2+B3 e in C4 quella della colonna C: =C2+C3.

In D4 riportiamo la somma della colonna D o della riga 4: =D2+D3 (oppure =B4+C4).

A questo punto possiamo riportare le formule per il calcolo dei test diagnostici.

A partire da D7, proseguendo verso il basso, scriviamo le seguenti formule:

- Sensibilità: =B2/B4
- Specificità: =C3/C4
- Valore predittivo positivo (VPP): =B2/D2
- Valore predittivo negativo (VPN): =C3/D3
- Indice di Youden (J): =(D7+D8)-1
- Probabilità pre-test: =B4/D4
- Likelihood ratio o rapporto di verosimiglianza positivo (LR+): =D7/(1-D8)
- Likelihood ratio o rapporto di verosimiglianza negativo (LR-): =(1-D7)/D8.

Inserendo i nostri dati nelle caselle B2, B3, C2 e C3 avremo automaticamente i nostri dati.

Ricordiamo che i veri positivi vanno inseriti in B2, i veri negativi in C3, i falsi positivi in C2 e i falsi negativi in B3.

Passiamo ora all'analisi dei dati per la terapia o per i fattori di rischio (fig. 2). Partendo da D18 e proseguendo verso il basso inseriamo in ordine le seguenti formule:

	A	B	C	D
1		M+	M-	Tot.
2	Esposti/test+	1	2	3
3	Non esposti/test-	4	3	7
4	Tot.	5	5	10
5				
6	Diagnosi			
7		Sensibilità	B2/B4	0,2
8		Specificità	C3/C4	0,6
9		VPP	B2/D2	0,333333
10		VPN	C3/D3	0,428571
11		J	(D7+D8)-1	-0,2
12		Pr. Pretest	B4/D4	0,5
13		LR+	D7/(1-D8)	0,5
14		LR-	(1-D7)/D8	1,333333
15				
16				
17				
18				
19				
20				
21				
22				

Fig. 1: prima parte della tabella 2 x 2 per il calcolo dei test diagnostici. In alto la griglia che costituisce la tabella 2 x 2

	A	B	C	D	E	F	G	H	I	J	K
1		M +	M -	Tot.							
2	Esposti /test+	1	2	3							
3	Non esposti/test -	4	3	7							
4	Tot.	5	5	10							
5											
16	Terapia										
17	o fattori di rischio										
18		EER	B202	0,333333							
19		CER	B303	0,571429							
20		RR	D18019	0,583333							
21		RRR	(D19-D18)/D19	0,416667							
22		ARR	D19-D18	0,238095							
23		NNT	1/D22	4,2							
24		OR	(B2xC3)	0,375							
25			(C2xB3)								
26		95%CI _{NNT}	1/ARR*(1,96*SE)		1,129473	-2,41392					
27			SE ²		0,109059						
28			SE		0,330242	4E27					
29			95%CI _{ARR}		0,885369						
30					-0,40918						
31											

Fig. 2: calcoli per i dati relativi alla terapia o ai fattori di rischio.

- Tasso di eventi nel gruppo trattato o esposto (EER): $=B2/D2$
- Tasso di eventi nel gruppo controllo o non esposto (CER) = $B3/D3$
- Rischio relativo (RR): $=D18/D19$
- Riduzione del rischio relativo: $=(D19-D18)/D19$
- Riduzione del rischio assoluto: $=D19-D18$
- Numero di pazienti da trattare (NNT): $=1/D22$
- Odds ratio (OR): $=(B2*C3)/(B3*C2)$

Per chi volesse divertirsi a calcolare l'intervallo di confidenza al 95% per il NNT prosegua come di sotto indicato.

Per prima cosa occorre ricavare l'errore standard (SE) e poi calcolare il 95%CI per l'ARR (vedi anche box 1 cap. 6).

Per il calcolo del SE² inserire in E27 la seguente formula:

$$=((D18*(1-D18))/D2)+((D19*(1-D19))/D3).$$

In E28 calcoliamo la radice quadrata del valore trovato in E27: $=RADQ(E27)$.

A questo punto in E29 e in E30 calcoliamo gli estremi del 95%CI di ARR:

in E29: $=D22+(1,96*E28)$:

in E30: $=D22-(1,96*E28)$.

A questo punto per ottenere gli estremi del 95%CI di NNT basta calcolare il reciproco dei due valori trovati.

Nella fig. 2 sono in E26 e F26, le formule sono rispettivamente le seguenti:

$$=1/E29; =1/E30.$$

Provate ora ad usare la vostra tabella 2 x 2 così costruita per verificare i dati che troverete negli articoli scientifici.

torgraf ■

Finito di stampare nel mese di Luglio 2010
presso lo stabilimento tipolitografico della **torgraf**
S.P. 362 km. 15,300 - Zona Industriale - 73013 **GALATINA** (Lecce)
Telefono +39 0836.561417 - Fax +39 0836.569901
e-mail: stampa@torgraf.it

