

Dalle conoscenze alle raccomandazioni Il metodo GRADE

Roberto Buzzetti
Epidemiologo, Rimini

Abstract

From literature to practice

Among the requirements of a good clinical guideline, grading has a considerable importance since not everything recommended has the same grading. For this purpose about twenty different grading systems have been developed. In order to unify the different grading systems at the beginning of 2000 a group of guidelines experts have defined a method that would maintain the link between the strength of the available evidence and the grade of the recommendation. The GRADE method confirms that the strength of recommendations has to be based on evidence levels but non only on them. Furthermore the grading system assigned basically on the study design can vary considering a formal assessment of methodological quality, quantity, consistency and applicability of the evidence, and also the cultural, epidemiological and normative aspects. In this way attention is given to study outcomes and not only to study design. For each important outcome quality assessment is based on the effectiveness of the single studies results. So great importance is given to the balance between pros and cons and the transferability of proofs within a specific setting.

Quaderni acp 2010; 17(5): 211-214

Key words Guidelines. EBM. GRADE. Working group. Evidence levels. Recommendations Strength

Tra i requisiti di una buona linea guida parecchia importanza ha il cosiddetto grading: non tutto ciò che si raccomanda di fare ha la stessa intensità. È stata sviluppata a questo proposito almeno una ventina di sistemi differenti di grading. Per tentare di unificare i vari sistemi, dall'inizio degli anni 2000 un gruppo di esperti sulle linee guida ha messo a punto un metodo per passare dai livelli di evidenza alla forza delle raccomandazioni. Il metodo GRADE conferma che la forza delle raccomandazioni deve essere basata sui livelli di evidenza, ma non deve limitarsi a questi. È possibile aumentare o diminuire il livello degli studi, inizialmente assegnato sulla base del disegno di studio, tenendo conto anche della qualità con cui gli studi sono condotti, della coerenza tra studi differenti e della diretta applicabilità dei risultati alla pratica, tenendo conto di fattori anche culturali, epidemiologici, organizzativi, normativi. L'attenzione, inoltre, si sposta dal disegno dello studio agli outcomes. La qualità delle prove viene valutata sulla validità dei risultati dei singoli studi, per ognuno degli outcomes importanti. Grande importanza assumono il bilancio tra pro e contro, e la trasferibilità delle prove nella pratica nello specifico contesto.

Parole chiave Linee guida. EBM. GRADE. Working Group. Livelli di evidenza. Raccomandazioni (forza delle)

Introduzione

L'EBM, le linee guida.

Conoscenze e raccomandazioni

Ormai da un paio di decenni siamo nell'era dell'EBM. Se questo fosse vero, dovremmo assistere a un "utilizzo coscienzioso (scrupoloso), esplicito (chiaro, preciso), giudizioso (prudente) delle migliori conoscenze disponibili al momento, nel processo decisionale riguardante la cura dei singoli pazienti" [1]. Prima di

cantar vittoria, dovremmo dunque verificare la tanto attesa "rivoluzione" almeno in due ambiti. Da una parte, il versante delle conoscenze; dall'altro, quello delle decisioni. Sul versante delle conoscenze sono stati fatti passi in avanti. L'aumentata coscienza critica, la grande rivoluzione (questa sì) tecnologica fanno sì che nel giro di poche ore, giorni o settimane tutto lo scibile su un tema di interesse sia a disposizione del professionista interessato al problema.

Sono inoltre venuti a compimento progetti di incredibile importanza sulla sintesi delle conoscenze.

La sfida lanciata dalla Cochrane Collaboration ci ha portato, nel giro di pochi anni, a un corpo di revisioni sistematiche assai ricco di rigore sia dal punto di vista quantitativo (box 1) sia soprattutto dal

BOX 1: CHE COSA CONTIENE LA COCHRANE LIBRARY

THE COCHRANE DATABASE OF SYSTEMATIC REVIEWS (CDRS): full-text delle revisioni sistematiche, regolarmente aggiornate (6162); revisioni 4247, di cui 30 nuove e 85 updates (22 nuove ricerche, - 6 con conclusioni modificate); protocolli delle revisioni in preparazione (1915), di cui 23 nuovi.

DATABASE OF ABSTRACTS OF REVIEWS OF EFFECTS (DARE): abstract strutturati di revisioni sistematiche non prodotte dalla Cochrane Collaboration e analizzate qualitativamente dal NHS Centre for Reviews and Dissemination di York (UK) (12.594).

THE COCHRANE CONTROLLED TRIALS REGISTER (CCTR/CENTRAL): registro di tutti i trial clinici identificati dai gruppi di revisione Cochrane allo scopo di preparare il proprio database specializzato (620.700). Ricerca sulle maggiori banche dati (Medline, Embase, Psycinfo, Cinahl). Ricerca manuale (dal 1945 in poi) su riviste, atti di congressi e su molte altre fonti non incluse in Medline o in altre banche dati.

THE COCHRANE METHODOLOGY REGISTER (CMR): raccolta bibliografica di articoli e testi sui metodi di conduzione delle revisioni sistematiche (13.164).

HEALTH TECHNOLOGY ASSESSMENT DATABASE (HTA): abstract di rapporti di Technology Assessment (8.273).

NHS ECONOMIC EVALUATION DATABASE (NHS EED): abstract di rapporti di valutazione economica (28.551).

Dati aggiornati al numero 5 del 2010 [2]

Per corrispondenza:
Roberto Buzzetti
e-mail: robuzze@gmail.com

leggere e fare

lato qualitativo, nella valutazione metodologica degli studi.

La seconda decisiva tappa porta dalle conoscenze alle raccomandazioni. Sono nate per questo le linee guida, all'inizio documenti confusi, talvolta simili a piccoli trattati o a protocolli, spesso basati sul consenso degli esperti. Pian piano si è fatta avanti una "scienza delle linee guida", e i numerosi studi sull'argomento hanno contribuito a fare chiarezza sui requisiti che una linea guida deve avere per poter essere considerata appunto una buona linea guida e non l'ennesimo atto di autorità di chi vuole imporre sugli altri le proprie pressioni e le proprie opinioni. Si è giunti a comprendere via via l'importanza del coinvolgimento delle parti in causa (necessità di un gruppo multidisciplinare), della necessità di basarsi sulle migliori evidenze, dell'importanza di un "grading" nelle raccomandazioni. Altri requisiti irrinunciabili sono la chiarezza degli obiettivi e delle linee guida stesse, l'indipendenza editoriale, l'applicabilità nella pratica quotidiana. È stato formalizzato uno strumento di valutazione, AGREE, che, esaminando le linee guida per ognuna di queste aree, permette di assegnare loro un punteggio di qualità [3].

Il problema del grading

Non tutto ciò che si raccomanda di fare o di evitare lo si raccomanda con la stessa intensità. Questo non solo nel mondo della medicina, ma in ogni campo della vita. Vi sono azioni importantissime da portare assolutamente a compimento o errori molto gravi da evitare assolutamente; ma vi sono anche suggerimenti, consigli per i quali sentiamo di spendere qualche energia, seppur in misura minore, forse perché ci sentiamo meno sicuri nel raccomandare. Vi sono infine decisioni sulle quali non ci sentiamo di sbilanciarci; vorremmo dire qualcosa in proposito, ma non abbiamo sufficienti prove per sostenere le ragioni del fare o del non fare.

Nel mondo delle linee guida è stata sviluppata almeno una ventina di sistemi differenti di grading. Qualcuno distingue tra "recommendation" e "option"; qualcun altro usa i verbi (*must, should, could, may...*), altri ancora risolvono il problema assegnando delle lettere o dei numeri per designare sia il livello di evidenza sia

BOX 2: LE SOCIETÀ SCIENTIFICHE E GLI ENTI CHE HANNO ADOTTATO O CHE UTILIZZANO IL METODO GRADE

Organizations that have endorsed or that are using GRADE [5]

WHO - Guidelines for World Health Organization Guidelines
 Endocrine Society Clinical Guidelines - USA
 ACCP American College of Chest Physicians Guidelines - USA
 UpToDate - Putting Clinical Information Into Practice - USA
 ASR Agenzia sanitaria regionale, Bologna - Italia
 Ministry of Health and Long-Term Care, Ontario - Canada
 Surviving Sepsis - International
 Ärztliches Zentrum für Qualität in der Medizin - Germany
 American Thoracic Society - USA
 American College of Physicians - USA
 The Cochrane Collaboration - International
 Kidney Disease: Improving Global Outcome - International
 European Society of Thoracic Surgeons - International
 Agency for Healthcare Research and Quality (AHRQ) - USA
 Society of Critical Care Medicine (SCCM) - USA
 National Institute for Clinical Excellence (NICE) - UK
 Norwegian Knowledge Centre for the Health Services - Norway
 The University of Pennsylvania Health System Center for Evidence-based Practice - USA
 German Center for Evidence-based Nursing "sapere aude" - Germany
 Evidence-based Nursing Südtirol, Alto Adige - Italy
 Society for Vascular Surgery - USA
 BMJ Clinical Evidence - UK
 EBM Guidelines - Finland/International
 Polish Institute for EBM - Poland
 European Respiratory Society (ERS) - Europe
 Japanese Society for Temporomandibular Joint - Japan
 National Board of Health and Welfare - Sweden
 COMPUS at The Canadian Agency for Drugs and Technologies in Health (CADTH) - Canada
 Infectious Diseases Society of America - USA
 Spanish Society for Family and Community Medicine
 Emergency Medical Services for Children National Resource Center - USA
 SBU - The Swedish Council on Technology Assessment in Health Care - Sweden
 The Scottish Intercollegiate Guidelines Network (SIGN) - UK
 Evidence-Based Tuberculosis Diagnosis (tbevidence.org) - Canada
 National & Gulf Center for Evidence Based Medicine (NGCEBM) - Saudi Arabia
 American Society for Gastrointestinal Endoscopy - USA
 European Association for the Study of the Liver - Europe
 CDC's Healthcare Infection Control Practices Advisory Committee (HICPAC) - USA
 Finnish Office for Health Technology Assessment - Finland
 NHS Quality Improvement Scotland - UK
 The American Association for the Study of Liver Diseases - USA
 The Canadian Cardiovascular Society - Canada
 La Société Canadienne de Cardiologie - Canada

la forza della raccomandazione. Va infatti tenuta ben presente questa distinzione fondamentale tra l'analisi asettica e il più possibile oggettiva delle conoscenze provenienti dagli studi e il mondo reale delle raccomandazioni che di fatto pretenderebbero di guidare le azioni dei professionisti.

Con questo spirito, all'inizio degli anni 2000, un gruppo di esperti sulle linee guida ha incominciato a riunirsi (gruppo GRADE) con vari obiettivi [4].

Il primo obiettivo era senz'altro quello di tentare di unificare i diversi sistemi di grading presenti al mondo. Il secondo obiettivo, ancora più interessante, trovare un metodo che permettesse di passare **dalle conoscenze (levels of evidence: "fino a che punto uno può confidare nel fatto che la stima dell'effetto sia corretta") alla forza delle raccomandazioni ("fino a che punto uno può confidare nel fatto che l'adesione alla raccomandazione darà più benefici che danni")**.

Questo non è banale. Si dice infatti che le raccomandazioni debbano essere basate sulle evidenze, cosa senz'altro vera: ma che dire di tutti quei sistemi che in passato assegnavano automaticamente e acriticamente la forza di una raccomandazione basandosi esclusivamente sul livello di evidenza? Studi randomizzati = livello I = forza A; studi non randomizzati = livello II = forza B; parere di esperti = livello III = forza C...

È evidente il limite di una classificazione così burocratica, e oltretutto anche indifendibile: se il livello degli studi che sostengono una raccomandazione determina totalmente la forza della raccomandazione stessa, che senso ha avere due scale di classificazione che in realtà sono una sola cosa?

Volendo sintetizzare qui al massimo il contributo del metodo GRADE, elenchiamo i principali aspetti innovativi. Nel **box 2** sono elencate le società scientifiche e gli enti che hanno aderito a questo metodo [5].

1. La forza delle raccomandazioni deve essere basata sui livelli di evidenza, ma non deve limitarsi a questi

Inizialmente viene assegnato un livello, con la seguente semplice scala:

- Studio randomizzato = alto
- Studio osservazionale = basso
- Altri tipi di prove = molto basso

... ma la novità è che poi il livello può essere ritoccato verso l'alto o verso il basso, tenendo conto anche di altri elementi. Si potrà dunque assistere a una *penalizzazione* di uno o anche due livelli, nei seguenti casi:

- **abbassare di un livello**

- gravi limitazioni alla qualità dello studio;
- incoerenza importante tra studi differenti;
- qualche dubbio sulla applicabilità;
- dati scarsi o imprecisi;
- alto rischio di "reporting bias";

- **abbassare di due livelli**

- molto gravi limitazioni alla qualità dello studio;
- importanti dubbi sulla applicabilità.

... o a una "promozione" di uno o due livelli, nei seguenti casi:

- **alzare di un livello**

- associazione **forte**:
 - rischio relativo >2 o <0,5
 - statisticamente significativa
 - in almeno due studi osservazionali
 - concordi tra loro e
 - senza plausibile rischio di confondimento;
- dimostrazione di un gradiente dose-risposta;
- possibilità che tutti i possibili confondenti abbiano ridotto l'effetto;

- **alzare di due livelli**

- associazione **molto forte**:
 - rischio relativo >5 o <0,2
 - statisticamente significativa
 - assenza di gravi difetti in grado di minacciare la validità degli studi.

Questo mette fine, in un certo senso, all'egemonia degli studi randomizzati: non basta l'abito (studio randomizzato) a fare il monaco (raccomandazione forte), perché uno studio randomizzato mal condotto, con piccoli numeri o altri difetti, verrà giudicato peggio di uno studio osservazionale ben fatto, con forte associazione. Si pensi agli studi sui rischi del fumo, altrimenti condannati sempre a essere "figli di un disegno minore": vi sono in proposito parecchi studi ovviamente non randomizzati, ma concordanti nel segnalare un rischio relativo di neoplasia polmonare molto elevato, un netto

gradiente dose-risposta; e che tengono conto del possibile confondimento legato a sesso, età, abitudini di vita, regione di residenza, alcol e altre variabili.

2. L'attenzione si sposta dal disegno dello studio agli outcome

La qualità delle prove viene valutata sulla validità dei risultati dei singoli studi, *per ognuno degli outcome importanti*. Il gruppo GRADE propone le fasi seguenti:

- identificare gli outcome importanti, critici per le decisioni (per esempio, nell'asma: FEV1, necessità di farmaci di sollievo della crisi, giorni di scuola persi, visite al Pronto Soccorso, effetti collaterali...);
- su ognuno di questi misurare la qualità globale delle prove disponibili;
- fare un bilancio oggettivo tra i benefici e i danni;
- assumere come livello globale di qualità delle prove la più bassa qualità delle prove trovata all'interno dei diversi outcome critici per la decisione.

Può, per esempio, accadere che un farmaco antiasmatico sia stato testato in modo randomizzato per quanto riguarda la variazione del FEV1 (livello di evidenza alto), ma che all'interno dello stesso studio o di altri studi l'effetto su un altro outcome, per esempio gli effetti collaterali, sia stato studiato senza randomizzazione (livello basso). Si concluderà per un livello di evidenza basso.

3. Le quattro componenti suggerite per una buona raccomandazione

La prima continua a essere, come detto, il disegno dello studio: meglio randomizzato che non randomizzato, meglio una coorte prospettica di uno studio retrospettivo, meglio la presenza che l'assenza del gruppo di controllo ecc.

La seconda è la qualità metodologica con cui lo studio è stato condotto. Nel caso degli studi randomizzati si utilizzano criteri come l'adeguatezza del nascondimento della sequenza di randomizzazione ("allocation concealment"), della cecità, del follow-up ecc.

La terza componente è la coerenza tra studi differenti.

La quarta infine è la cosiddetta "directness" che consiste nella misura in cui soggetti, interventi e outcome sono simili

li a quelli di riferimento. Per esempio, potrebbe esservi incertezza sulla directness delle prove se i soggetti di riferimento fossero di età differente o affetti da maggior co-morbilità rispetto a quelli degli studi. Per alcune terapie, per esempio gli interventi comportamentali in cui è probabile che le differenze culturali siano importanti, la directness è ovviamente minore; lo stesso dicasi per farmaci differenti da quelli degli studi, ma appartenenti alla stessa classe.

In sintesi, e in pratica, coloro che producono una raccomandazione dovrebbero considerare quattro fattori principali:

- i bilanci tra pro e contro, tenendo conto della dimensione stimata dell'effetto per gli outcome principali, dei limiti di confidenza attorno a quelle stime, e del valore relativo per ogni outcome;
- la qualità delle prove (*vedi sopra*);
- la trasferibilità delle prove nella pratica in uno specifico contesto, tenendo conto dei fattori importanti in grado di modificare le dimensioni degli effetti attesi, come per esempio la disponibilità nelle vicinanze di un ospedale o della expertise necessaria;
- l'accettazione da parte dei pazienti di certi trattamenti; la presenza di barriere culturali, organizzative, normative l'incertezza sul rischio di base per la popolazione di interesse. È noto come spesso gli studi utilizzino pazienti a rischio elevato (esempio: soggetti affetti da reflusso gastro-esofageo importante) mentre l'applicazione è poi spesso diretta a pazienti a rischio più modesto.

Nel formulare le raccomandazioni sarebbe opportuno utilizzare uno schema semplice, come il seguente, su due soli livelli:

- **DA FARE** ("do it") o **DA NON FARE** ("don't do it"): indica un giudizio che verrebbe dato dalla maggior parte delle persone correttamente informate;
- **PROBABILMENTE DA FARE** ("probably do it") o **PROBABILMENTE DA NON FARE** ("probably don't do it"): indica un giudizio che verrebbe dato dalla maggior parte delle persone correttamente informate, ma non da una minoranza importante.

Conclusioni

Il metodo GRADE, in conclusione, rappresenta un binario che, se seguito dai più, dovrebbe portarci a raccomandazioni sempre più credibili e meno rigide che in passato, anche se questa minor rigidità non corrisponde a una maggiore lassità ma anzi a un rigore ancora maggiore. Per dirla con i suoi promotori, "gli utilizzatori di linee guida per la pratica clinica e di altre raccomandazioni hanno bisogno di sapere quanta fiducia possano riporre nelle raccomandazioni. L'utilizzo di metodi sistematici ed espliciti nella costruzione di criteri può ridurre gli errori e migliorare la comunicazione". ♦

Bibliografia

- [1] Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. *JAMA* 1992;268:2420-5.
- [2] <http://www.thecochranelibrary.com/view/0/AboutTheCochraneLibrary.html>.
- [3] http://asr.regione.emilia-romagna.it/wcm/asr/collana_dossier/doss060/link/doss60.pdf.
- [4] GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490 (19 June), doi:10.1136/bmj.328.7454.1490.
- [5] <http://www.gradeworkinggroup.org/society/index.htm>.

SECONDA EDIZIONE DEL PREMIO NAZIONALE "NATI PER LEGGERE"

Dopo il successo con cui è stata accolta la prima edizione, riparte il Premio nazionale "Nati per Leggere" (bando su www.natiperleggere.it) che sostiene i migliori libri, progetti editoriali e progetti di promozione alla lettura per i più piccoli. È rivolto a editori, bibliotecari, insegnanti, librai, educatori, pediatri ed enti locali.

Le sezioni

- Sezione **Libri in cantiere**: premia il miglior progetto editoriale inedito.
- Sezione **Nascere con i libri**: premia il miglior libro per bambini di età tra 6 e 36 mesi.
- Sezione **Crescere con i libri**: segnala il miglior libro per bambini di età compresa tra 3 e 6 anni, pubblicato in Italia tra il 2005 e il 2010 e relativo al tema "Libri per ridere: l'umorismo nei libri per bambini".
- Sezione **Reti di libri**: premia il miglior progetto di promozione della lettura rivolto ai bambini in età compresa tra 0 e 5 anni, sviluppato secondo le linee di "Nati per Leggere", che coinvolga il più ampio numero di genitori, bibliotecari, pediatri, insegnanti, educatori, volontari, ecc. attraverso un lavoro di rete.
- Sezione **Pasquale Causa**: premia il pediatra che, aderendo al progetto Nati per Leggere, promuova nel modo più efficace la pratica della lettura ad alta voce in famiglia.

I premi

Libri in cantiere elargirà un contributo di 2000 euro per la pubblicazione del volume.

Nascere con i libri e **Crescere con i libri** assicureranno l'acquisto del volume vincitore per un importo di 2000 euro e garantirà la distribuzione del libro a pediatri, biblioteche, nidi e scuole dell'infanzia.

Reti di libri e **Pasquale Causa** elargiranno 2000 euro da impegnare nell'acquisto di libri nell'edizione speciale "Nati per Leggere".

**Scadenza per la partecipazione
31 gennaio 2011**

Informazioni

Segreteria Organizzativa:
npl-premio@aib.it,
zacheo@salonelibro.it
tel. 011 5184268 int. 954