

Keep it simpler. Proposta per un nuovo sistema di valutazione delle linee-guida

Le linee-guida sono raccomandazioni sviluppate in modo sistematico per assistere operatori sanitari e pazienti nelle decisioni sulla gestione appropriata di specifiche condizioni cliniche. Definire il grado di validità delle informazioni scientifiche poste alla base di una serie di raccomandazioni e chiarire quanta importanza si attribuisce alla loro applicazione costituiscono uno dei caratteri distintivi del processo di produzione di linee-guida basate sulle prove di efficacia.

La diffusione di sistemi diversi di valutazione delle linee-guida indebolisce l'esercizio di valutazione; sono necessari criteri di valutazione semplici, trasparenti, metodologicamente chiari che descrivano in maniera limpida la qualità dell'evidenza e la forza delle indicazioni alla base. L'American College of Chest Physicians (ACCP) ha costituito un gruppo di lavoro per concordare un sistema di valutazione da utilizzare in tutte le sue linee-guida; l'analisi condotta è stata pubblicata su Chest. Pubblichiamo di seguito l'intervista a Holger J. Schünemann, uno degli autori dell'articolo, nonché i riassunti in italiano e in inglese del lavoro originale.*



**Intervista a
Holger J. Schünemann***

Condividere conoscenze non è facile, viste: 1) la mancanza di un accordo generale su ciò che definisce un' "evidenza di alta qualità", e 2) la complessità dei sistemi di valutazione applicati. Quali sono i principali obiettivi del vostro lavoro?

Lo sforzo va ben oltre l'ACCP. L'opera della task force dell'ACCP per valutare la qualità dell'evidenza è un esempio di come un lavoro su più vasta scala si sia cominciato a diffondere e abbia trovato consensi. Il lavoro è strettamente legato al gruppo del GRADE (www.gradeworkinggroup.org), nato dalla collaborazione internazionale di esperti di produzione e metodologia delle linee-guida (LG) e il cui scopo è quello di elaborare LG migliori. L'obiettivo principale del gruppo di lavoro del GRADE è quello di rendere le LG e le loro indicazioni trasparenti ed esplicite per gli utenti, ossia i fruitori di LG.

A tale scopo è importante sviluppare e diffondere un sistema di valutazione che comunichi in maniera chiara la qualità dell'evidenza e la forza delle indicazioni che sono alla base e che devono poter essere universalmente riconoscibili. La validazione di questo tipo di sistema, secondo l'ACCP, dovrebbe provenire senza dubbio da una società scientifica. Pertanto, potremmo convincere la task force dell'ACCP ad adottare la metodologia del GRADE. Il mio coinvolgimento con l'ACCP è il risultato di molti anni di collaborazione con loro nel settore dell'elaborazione di LG in qualità di uno dei quattro curatori di una delle LG più diffusamente conosciute: le LG dell'ACCP sulla terapia antitrombotica e trombolitica basata sulle evidenze.

Il gruppo di lavoro ha deciso di classificare le indicazioni in due soli livelli, forte e debole: può spiegarci la ragione di questa scelta?

I medici, primi fruitori di LG, hanno poco tempo per reperire risposte su interrogativi che raramente

si trovano ad affrontare. Un sistema di valutazione a due livelli ha il merito della semplicità, qualità molto richiesta dal clinico. Inoltre due livelli facilitano, da parte del clinico, l'interpretazione chiara di ciò che le indicazioni forti e deboli implicano. Tale approccio, inoltre, ci consente di capire il grado di certezza con cui viene data un'indicazione sicura o non sicura. "Sicuro" significa che i benefici superano i rischi (o viceversa quando vengono prodotte indicazioni "contro" il fare qualche cosa) e "debole" significa che non si è certi e che pertanto i clinici devono approfondire la questione, molto più attentamente di quanto non lo debbano fare per le indicazioni forti.

L'approccio evidence-based alla pratica medica prende in considerazione la prospettiva del paziente, le sue esperienze e preferenze. Voi ne avete tenuto conto?

Sì, i clinici sono sempre più consapevoli dell'importanza dei valori e delle preferenze del paziente nel decision making clinico. Anche il sistema di valutazione del GRADE riconosce esplicitamente tale importanza nel processo di decision making. Un modo di interpretare le indicazioni forti e deboli è proprio in relazione ai valori e alle preferenze

*Holger J. Schünemann è professore associato presso il Department of Medicine dell'Università di Buffalo (NY), presso il Department of Clinical Epidemiology and Biostatistics della McMaster University, Canada, e responsabile dell'Unità di Sviluppo della ricerca clinica e trasferimento dell'INFORMAZIONE (INFORMA) dell'Istituto Nazionale Tumori Regina Elena di Roma. È membro del GRADE e document editor per l'American Thoracic Society (ATS).

del paziente. Per decisioni nelle quali è chiaro che i benefici superano di gran lunga i danni e i rischi, al contrario, i danni e i rischi vanno ben al di là dei benefici, tutti i pazienti virtualmente faranno la stessa scelta. In questo caso, gli esperti di LG possono suggerire un'indicazione forte. All'opposto, ci sono scelte nelle quali i valori e le preferenze del paziente avranno un ruolo determinante e nelle quali i pazienti, come risultato, faranno scelte differenti. Quando, trasversalmente alla loro gamma di valori, pazienti pienamente informati sono esposti a più scelte, gli esperti di LG dovrebbero sottolineare la "debolezza" delle indicazioni suggerite.

Clinici, infermieri e manager della salute sono abituati alla comune gerarchia delle evidenze basata sugli studi controllati randomizzati (RCT) al vertice, e gli studi osservazionali alla base. Il gruppo di lavoro del GRADE propone un approccio abbastanza innovativo; può descrivercelo?

I primi sistemi di valutazione metodologica della qualità si rifacevano in primo luogo al disegno di base dello studio (per esempio gli RCT o gli studi osservazionali). Il disegno di studio rimane di fondamentale importanza nel determinare l'attendibilità della stima degli effetti benefici e dannosi di un trattamento. Per le differenze prognostiche tra i gruppi e la mancanza di garanzie – come la cecità – che consentono di annullare quei fattori che possono condizionare gli outcome, l'evidenza basata sugli studi osservazionali sarà, in generale, sensibilmente più debole di quella basata sugli RCT. Gli ultimi anni hanno testimoniato, comunque, una consapevolezza maggiore di fattori ulteriori che influenzano l'attendibilità delle stime di rischi e di benefici.

Le indicazioni basate sul sistema del GRADE utilizzeranno d'ora in

avanti un sistema articolato in quattro categorie di qualità dell'evidenza: alta, moderata, bassa e molto bassa. La task force dell'ACCP ha deciso di adottare un approccio leggermente diverso di sole tre categorie (alta, moderata e bassa), in quanto i quesiti clinici con cui l'ACCP ha a che fare sono spesso supportati da evidenza di qualità migliore (rispetto, ad esempio, ad altri di sanità pubblica e specialità chirurgiche nei quali la qualità dell'evidenza è spesso più modesta).

In generale, gli esperti di LG dovrebbero avere a disposizione revisioni sistematiche dell'evidenza (ossia basate su una metodologia standardizzata) per quanto riguarda i benefici e i rischi delle strategie prese in considerazione. Avremo la più forte evidenza quando tali revisioni riveleranno uno o più RCT ben disegnati e ben eseguiti, da cui risultati coerenti direttamente applicabili. L'evidenza di alta qualità può inoltre provenire, in circostanze inusuali, da studi osservazionali ben fatti (per esempio studi di coorte ben condotti e controllati) da cui possono derivare effetti molto ampi degni di considerazione.

RCT con importanti limitazioni e studi osservazionali rigorosi con conseguenti risultati di una certa portata costituiscono la categoria di qualità moderata. Studi osservazionali ben fatti e, occasionalmente, RCT con limitazioni molto gravi verranno stimati come evidenza di qualità bassa. La categoria di qualità molto bassa include studi osservazionali scarsamente controllati e osservazioni cliniche non sistematiche (per esempio sequenze di casi o casi clinici). Tale valutazione segue il principio secondo cui tutti gli studi clinici rilevanti e le osservazioni forniscono evidenza, la cui qualità varia. In ogni caso, il sistema esplicita il fatto che l'opinione dell'esperto non è una categoria di evidenza. Essa rappresenta un'interpretazione dell'evidenza, inclusa quella che va

dalle osservazioni nella pratica di un esperto (osservazioni non controllate) all'interpretazione degli RCT o metanalisi di cui sia a conoscenza l'esperto. Questa categorizzazione si basa sul principio secondo cui qualsiasi studio clinico rilevante fornisce evidenza, la cui qualità varia.

Cosa fare quando la qualità dell'evidenza cambia da un outcome all'altro?

I sistemi di valutazione precedenti non prendevano in considerazione uno dei più importanti aspetti della formulazione di un quesito clinico. Ossia che le decisioni cliniche implicano sempre un trade-off tra benefici e rischi. Pertanto, porre un quesito dovrebbe implicare, a valle, la considerazione di tutti gli outcome e la valutazione di tutti i possibili esiti dell'intervento, inclusi quelli usualmente considerati dannosi. La qualità dell'evidenza cambierà spesso tra efficacia primaria e outcome di tossicità; frequentemente tra outcome di efficacia e costi; quasi sempre tra outcome di efficacia ed effetti collaterali gravi anche se rari. In molti casi, gli outcome di efficacia saranno i più importanti ed evidenze di danno (se rare) non saranno determinanti per il decision making; in questi casi gli esperti di LG possono basare la valutazione della qualità dell'evidenza esclusivamente su questi endpoint. Gli esperti dovrebbero, comunque, valutare se gli endpoint di tossicità sono cruciali per la decisione relativa alla terapia ottimale. Se lo sono, allora il panel dovrebbe tenere conto della qualità dell'evidenza di quegli endpoint, e dovrebbe arrivare ad una valutazione finale circa la qualità dell'evidenza fissando, come qualità globale, quella più bassa tra tutti gli outcome critici.

Facciamo un esempio. Poniamo che, in un RCT, l'efficacia di un nuovo intervento venga misurata con test da sforzo, e venga mostrato

un lieve miglioramento in termini di distanza percorsa nella camminata: ne risulterebbe un'evidenza di qualità alta. Ma se sussistessero gravi preoccupazioni in termini di sicurezza, associate per esempio ad un incremento di rischio miocardico, che non sono valutabili nell'RCT ma che vengono invece messe in evidenza in uno studio osservazionale (qualità bassa), allora l'evidenza globale dovrebbe essere considerata bassa.

Per concludere, quali sono le prospettive per il nuovo sistema di valutazione proposto delle LG proposto dal GRADE?

L'impegno del GRADE – condiviso dall'ACCP attraverso la sua task

force – è quello di cambiare il mondo della valutazione della qualità di evidenza e di forza delle indicazioni. La Cochrane Collaboration si sta muovendo verso l'adozione dell'approccio del GRADE per la valutazione della qualità metodologica, e anche il QUORUM (Quality of Reporting of Systemic Reviews) sta per adottare l'approccio del GRADE approvandone le raccomandazioni all'interno del suo statement revisionato. La Endocrine Society è stata la prima organizzazione nord americana ad aderire al GRADE per le sue indicazioni, mentre un'altra importante organizzazione, la UpToDate®, per il primo manuale medico elettronico, ha adottato una ver-

sione leggermente modificata del GRADE. Altre organizzazioni, quali l'American Thoracic Society e il BMJ Publishing Group, editore di Clinical Evidence, stanno per verificare il possibile utilizzo del metodo del GRADE. Per finire, e di grande rilievo, la World Health Organization ha ufficialmente adottato il GRADE e lo applicherà ad un lavoro in corso di LG sul trattamento e la prevenzione dell'influenza aviaria. C'è, pertanto, la reale possibilità che il GRADE possa realizzare per le indicazioni sulla sanità qualcosa di simile al sistema universalmente riconosciuto del semaforo per cui la luce rossa indica "stop" e quella verde "avanti". **bif**

Riassunto

Guyatt G, Gutterman D, Baumann MH, Addrizzo-Harris D, Hylek EM, Phillips B, Raskob G, Lewis SZ, Schünemann JH. Grading strength of recommendations and quality of evidence in clinical guidelines: report from an American College of Chest Physicians task force. Chest 2006; 129: 174-81.

Mentre la valutazione dell'efficacia delle raccomandazioni e della qualità delle evidenze rafforza l'utilità delle linee-guida cliniche, la diffusione di sistemi diversi di valutazione delle linee-guida indebolisce il valore dell'esercizio di valutazione. Una task force dell'American College of Chest Physicians (ACCP) ha formulato i criteri del sistema di valutazione da utilizzare in tutte le linee-guida ACCP. Tali criteri prevedono semplicità e trasparenza, chiarezza della metodologia, e coerenza del procedimento di valutazione con gli approcci metodologici attuali. Il gruppo di lavoro ha esaminato i sistemi attualmente disponibili, e infine ha modificato un approccio formulato dal gruppo internazionale GRADE. Lo schema di valutazione classifica le raccomandazioni come forti (grado 1) o deboli (grado 2), a seconda del rapporto di benefici, rischi, responsabilità, eventualmente costi, e il grado di confidenza nella stima di benefici, rischi, e responsabilità. Il sistema classifica la qualità dell'evidenza come alta (grado A), moderata (grado B), o bassa (grado C) a seconda di fattori quali disegno dello studio, coerenza dei risultati, chiarezza dell'evidenza. Per tutte le sue linee-guida, l'ACCP ha adottato un approccio semplice e tra-

sparente per la valutazione delle raccomandazioni, il quale è coerente con gli attuali sviluppi in materia. Il trend nell'uniformità degli approcci di valutazione rafforzerà l'utilità delle linee-guida per la pratica dei medici.

Abstract

While grading the strength of recommendations and the quality of underlying evidence enhances the usefulness of clinical guidelines, the profusion of guideline grading systems undermines the value of the grading exercise. An American College of Chest Physicians (ACCP) task force formulated the criteria for a grading system to be utilized in all ACCP guidelines that included simplicity and transparency, explicitness of methodology, and consistency with current methodological approaches to the grading process. The working group examined currently available systems, and ultimately modified an approach formulated by the international GRADE group. The grading scheme classifies recommendations as strong (grade 1) or weak (grade 2), according to the balance among benefits, risks, burdens, and possibly costs, and the degree of confidence in estimates of benefits, risks, and burdens. The system classifies quality of evidence as high (grade A), moderate (grade B), or low (grade C) according to factors that include the study design, the consistency of the results, and the directness of the evidence. For all future his guidelines, the College has adopted a simple, transparent approach to grading recommendations that is consistent with current developments in the field. The trend toward uniformity of approaches to grading will enhance the usefulness of practice guidelines for clinicians.