evidence
open access journal published by the GIMBE Foundation

Standards & Guidelines

OPEN ACCESS

QUADAS-2: strumento per valutare la qualità degli studi di accuratezza diagnostica

Penny F. Whiting[1*], Anne W.S. Rutjes[2], Marie E. Westwood[3], Susan Mallett[4], Jonathan J. Deeks[5], Johannes B. Reitsma[6], Mariska M.G. Leeflang[7], Jonathan A.C. Sterne[1], Patrick M.M. Bossuyt[7] and the QUADAS-2 Group

[1]School of Social and Community Medicine, University of Bristol, [2]Division of Clinical Epidemiology and Biostatistics, Institute of Social and Preventive Medicine, University of Bern, [3]Kleijnen Systematic Reviews, [4]Centre for Statistics in Medicine and Department of Primary Health Care, Wolfson College Annexe, Univeristy of Oxford, [5]Unit of Public Health, Epidemiology & Biostatistics, University of Birmingham, [6]Julius Center for Health Sciences and Primary Care, UMC Utrecht, [7]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, AMC, University of Amsterdam

ABSTRACT

Nel 2003 è stato sviluppato lo strumento QUADAS per le revisioni sistematiche degli studi di accuratezza diagnostica. QUADAS-2 ne rappresenta l'aggiornamento realizzato sulla base dell'esperienza, di report aneddotici e feedback che hanno suggerito aree di miglioramento. QUADAS-2 è strutturato in 4 domini: selezione dei pazienti, test in studio, standard di riferimento, flusso e timing. Per ogni dominio viene valutato il rischio di bias e per i primi 3 anche l'applicabilità. Sono inclusi quesiti guida per orientare il giudizio sul rischio di bias.

Lo strumento QUADAS-2 prevede 4 fasi: riassumere il quesito della revisione sistematica, adattare lo strumento e produrre istruzioni specifiche per la revisione sistematica, costruire un diagramma di flusso per lo studio primario e valutare bias e applicabilità. QUADAS-2 garantirà maggiore trasparenza nella classificazione dei bias e dell'applicabilità degli studi primari di accuratezza diagnostica.

Citazione. Whiting PF, Rutjes AWS, Westwood ME et al. QUADAS-2: strumento per valutare la qualità degli studi di accuratezza diagnostica. Evidence 2016;8(2): e1000131.

Pubblicato 2 febbraio 2016

Fonti di finanziamento. Questo studio è stato finanziato dal Medical Research Council nell'ambito del Medical Research Council–National Institute for Health Research Methodology Research Programme e dal National Institute for Health Research. Il dott. Mallett è finanziato dal Cancer Research UK. Il dott. Leeflang è finanziato dalla Netherlands Organization for Scientific Research (project 916.10.034).

Conflitti di interesse. Le disclosures sono disponibili a: www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M11-1238.

Provenienza. Tradotto con permesso da: Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529-36. Disponibile a: http://annals.org/article.aspx?articleid=474994 (ultimo accesso 2 febbraio 2016).

*E-mail: penny.whiting@bristol.ac.uk.

evidence
open access journal published by the GIMBE Foundation

Standards & Guidelines OPEN ACCESS

Le revisioni sistematiche (RS) degli studi di accuratezza diagnostica sono spesso caratterizzate da risultati notevolmente eterogenei conseguenti alle differenze di pianificazione e conduzione degli studi inclusi, dei quali è fondamentale valutare la qualità. Dalla sua pubblicazione nel 2003, lo strumento QUADAS (*Quality Assessment of Diagnostic Accuracy Studies*) è stato ampiamente utilizzato[1,2]. Oltre 200 abstract inseriti nel *Database of Abstracts of Reviews of Effects* menzionano questo strumento, citato più di 500 volte. L'*Agency for Healthcare Research and Quality* (AHRQ), la *Cochrane Collaboration*[3] e il *National Institute for Health and Clinical Excellence* (NICE) ne raccomandano l'utilizzo per condurre RS degli studi di accuratezza diagnostica.

Lo strumento QUADAS originale comprende 14 item che valutano il rischio di bias, le fonti di variabilità (applicabilità) e la qualità del reporting: per ogni item occorre indicare "sì", "no" o "non chiaro"[1]. Grazie a nuove evidenze, esperienza degli autori, report degli utilizzatori e feedback della *Cochrane Collaboration* sono state suggerite aree di potenziale miglioramento. In particolare, gli utilizzatori del QUADAS hanno riportato criticità relative alla valutazione di alcuni item (es. spettro dei pazienti, risultati dei test non interpretabili o intermedi e whithdrawals); alla possibile sovrapposizione tra item (es. bias di verifica parziale e *whithdrawals*); alle situazioni in cui QUADAS è difficile da utilizzare (es. quando lo standard di riferimento richiede il follow-up dei pazienti).

Questo articolo descrive QUADAS-2, versione aggiornata dello strumento.

METODI

Lo sviluppo di QUADAS-2 è basato sull'approccio a 4 step proposto da Moher et coll.[4]: definizione dell'obiettivo, revisione delle evidenze, meeting di consenso e test pilota al fine di perfezionare lo strumento.

Definizione dell'obiettivo. Un comitato direttivo di 9 esperti nel campo della ricerca diagnostica, molti dei quali avevano partecipato allo sviluppo dello strumento originale, ha definito le caratteristiche principali di QUADAS-2. Innanzitutto, il concetto di "qualità" è stato separato in due dimensioni: "rischio di bias" e "problemi di applicabilità". La qualità degli studi di accuratezza diagnostica è stata definita sia in termini di rischio di bias che di problemi di applicabilità di uno studio, valutando in particolare se: 1) le stime di accuratezza diagnostica hanno evitato il rischio di bias; 2) gli studi primari sono applicabili al quesito della RS. Il bias si verifica quando errori sistematici nel disegno o nella conduzione di uno studio ne distorcono i risultati. Uno studio primario può avere un'applicabilità limitata alla popolazione in studio se, rispetto al quesito definito dalla RS, arruola partecipanti con differenti caratteristiche demografiche o cliniche, se il test in studio viene applicato o interpretato in modo diverso o se la definizione della condizione target è differente.

Nel QUADAS-2, la riduzione del numero di domini principali è finalizzata a minimizzare le sovrapposizioni e ad estendere lo strumento alla valutazione sia di studi comparativi tra diversi test diagnostici, sia di studi che prevedono standard di riferimento basati sul follow-up, fatta espressa esclusione degli studi che valutano fattori prognostici. È stato inoltre proposto di modificare la valutazione sì/no/non chiaro del QUADAS originale in "basso rischio di bias" o "elevato rischio di bias", analogamente allo strumento sviluppato dalla Cochrane Collaboration per valutare il rischio di bias nei trial clinici. Infatti, un giudizio esplicito sul rischio di bias è più informativo e i feed-back relativi allo strumento Cochrane per classificare il rischio di bias hanno suggerito che la classificazione sì/no/non chiaro era fuorviante[5].

Revisione delle evidenze. Durante lo sviluppo di QUADAS-2 sono state condotte 4 revisioni[6]. La prima ha indagato le modalità di valutare e integrare la qualità in 54 RS di accuratezza diagnostica pubblicate tra il 2007 e il 2009. La seconda revisione ha utilizzato un questionario web-based per raccogliere un feedback strutturato da 64 autori di RS che avevano utilizzato QUADAS. La terza revisione condotta su 101 studi ha permesso di aggiornare le fonti di bias e di variabilità negli studi di accuratezza diagnostica[7]. La quarta revisione ha esaminato 8 studi che hanno valutato QUADAS: i dati completi saranno oggetto di un'ulteriore pubblicazione. Le evidenze emerse da queste revisioni hanno informato le decisioni sugli argomenti oggetto di discussione nel meeting di consenso. Sono stati sintetizzati i problemi riportati con lo strumento QUADAS originale, le evidenze per ciascun item originale e definiti i possibili item aggiuntivi relativi a bias e applicabilità. È stato inoltre redatto un elenco di ulteriori item candidati per la valutazione di studi di confronto tra differenti test in studio.

Meeting di consenso. Il 21 settembre 2010 a Birmingham (UK) si è svolto il meeting per sviluppare la prima bozza di QUADAS-2. I 24 partecipanti del gruppo QUADAS-2 erano esperti metodologi e revisori che lavorano sulle RS di accuratezza diagnostica. Sono state presentate sintesi delle evidenze scientifiche e i partecipanti sono stati divisi in piccoli gruppi per discutere dei contenuti dello strumento (protocollo del test, procedure di verifica, interpretazione, analisi, selezione dei pazienti o disegno dello studio, item di valutazione comparativa tra test), dell'applicabilità e delle decisioni concettuali. Sulla base dei risultati concordati nel meeting, i membri del comitato direttivo hanno prodotto la prima bozza di QUADAS-2.

Test pilota e perfezionamento. Successivamente lo strumento QUADAS-2 è stato perfezionato attraverso vari round di test pilota. Ad ogni round erano previsti questionari online per raccogliere feedback strutturati, ma

evidence
open access journal published by the GIMBE Foundation
Standards & Guidelines
OPEN ACCESS

sono stati accettati anche feedback verbali o via e-mail. I test pilota sono stati condotti dai membri del gruppo QUADAS-2, dai partecipanti al *Cochrane Colloquium* a Keystone (Colorado, ottobre 2010), da esperti di RS presenti al tavolo tecnico del NICE e da studenti svizzeri in scienze biomediche.

Il QUADAS-2 è stato sperimentato da coppie di revisori in 5 RS relative a diversi argomenti. La riproducibilità tra osservatori variava in maniera considerevole e c'era maggiore accordo sull'applicabilità che sul rischio di bias. Un'ulteriore coppia di revisori esperti ha testato lo strumento su una RS con test in studio multipli. Il feedback di questi revisori ha mostrato una scarsa riproducibilità tra osservatori e alcuni problemi nell'applicazione del dominio sull'accuratezza comparativa degli studi.

Tenendo conto di queste problematiche e della limitate evidenze sul rischio di bias e sulle fonti di variabilità in questi studi, QUADAS-2 non prevede criteri per valutare gli studi che confrontano test multipli. Il feedback relativo a tutte le altre fasi del processo è stato positivo: in particolare, tutti i partecipanti hanno preferito QUADAS-2 allo strumento originale.

Ruolo dei finanziatori. Questo articolo è stato finanziato da: *Medical Research Council, National Institute for Health Research*, *Cancer Research UK* e *Netherlands Organization for Scientific Research* (916.10.034). Gli sponsor non hanno avuto alcun ruolo nel disegno dello studio, nella raccolta, analisi e interpretazione dei dati, nella stesura del report o nella decisione di sottomettere il manoscritto per la pubblicazione.

QUADAS-2

Lo strumento QUADAS-2 è stato progettato per valutare la qualità degli studi di accuratezza diagnostica e dovrebbe essere utilizzato in aggiunta all'estrazione dei dati (es. disegno dello studio, risultati, etc.) da utilizzare nella RS. QUADAS-2 è strutturato in 4 domini: selezione dei pazienti, test in studio, standard di riferimento e flusso dei pazienti e timing di test in studio e standard di riferimento (tabella 1).

L'utilizzo dello strumento prevede quattro fasi: 1) esplicitare il quesito della RS; 2) definire precise istruzioni per la RS; 3) rivedere il diagramma di flusso pubblicato nello studio primario o costruirlo se non riportato; 4) valutare bias e applicabilità. Ogni dominio viene valutato in termini di rischio di bias e i primi 3 anche in relazione all'applicabilità. Per aiutare a valutare il rischio di bias, QUADAS-2 fornisce quesiti guida correlati ai potenziali bias dello studio.

FASE 1. QUESITO DELLA RS

I revisori devono anzitutto descrivere il quesito della RS in termini di pazienti, test in studio, standard di riferimento e condizione target. Considerato che l'accuratezza

Tabella 1. Rischio di bias e giudizio di applicabilità nel QUADAS-2

Dominio	Selezione dei pazienti	Test in studio	Standard di riferimento	Flusso e timing
Descrizione	Descrivere i metodi di selezione dei pazienti Descrivere i pazienti inclusi: test precedenti, presentazione clinica, uso previsto del test in studio e setting	Descrivere il test in studio e le modalità di somministrazione e interpretazione	Descrivere lo standard di riferimento e le modalità di somministrazione e interpretazione	Descrivere tutti i pazienti che non ricevono il test in studio o lo standard di riferimento o che sono stati esclusi dalla tabella 2 x 2 (si veda il diagramma di flusso) Descrivere l'intervallo tra il test in studio e lo standard di riferimento, oltre a qualunque intervento somministrato
Quesiti guida (*sì, no, non chiaro*)	È stato arruolato un campione di pazienti consecutivo o casuale? È stato evitato il disegno di studio caso-controllo? Lo studio ha evitato esclusioni di pazienti inappropriate?	I risultati del test in studio sono stati interpretati senza conoscere i risultati dello standard di riferimento? Il valore soglia eventualmente utilizzato era predefinito?	Lo standard di riferimento è adeguato per classificare correttamente la condizione target? I risultati dello standard di riferimento sono stati interpretati senza conoscere i risultati del test in studio?	L'intervallo temporale tra l'esecuzione del test in studio e dello standard di riferimento è adeguato? Tutti i pazienti hanno ricevuto lo stesso standard di riferimento? Tutti i pazienti arruolati sono stati inclusi nell'analisi?
Rischio di bias (*elevato, basso, non chiaro*)	La selezione dei pazienti potrebbe essere fonte di bias?	La somministrazione e l'interpretazione del test in studio potrebbe essere fonte di bias?	La somministrazione e l'interpretazione dello standard di riferimento potrebbe essere fonte di bias?	Il flusso dei pazienti potrebbe essere fonte di bias?
Rischio di problemi di applicabilità (*elevato, basso, non chiaro*)	I pazienti inclusi e il setting di arruolamento potrebbero non corrispondere al quesito della RS?	Il test in studio, la sua esecuzione o l'interpretazione potrebbero differire dal quesito della RS?	La condizione target definita dallo standard di riferimento potrebbe non corrispondere al quesito della RS?	-

evidence
open access journal published by the GIMBE Foundation
Standards & Guidelines
OPEN ACCESS

di un test può dipendere dal setting in cui sarà utilizzato nel percorso diagnostico, i revisori devono inoltre descrivere i pazienti in termini di setting, l'uso previsto del test in studio, le caratteristiche dei pazienti e l'esecuzione di eventuali test precedenti[8,9].

FASE 2. ADATTAMENTO SPECIFICO ALLA REVISIONE

È fondamentale adattare QUADAS-2 a ciascuna RS, aggiungendo o eliminando quesiti guida e sviluppando precise istruzioni per valutare ciascun quesito guida e usare questa informazione per stimare il rischio di bias (figura 1). Come primo step occorre considerare se uno o più quesiti guida non possono essere applicati alla RS o se specifiche caratteristiche della RS non sono adeguatamente affrontati dai quesiti guida. Ad esempio, se una RS valuta l'accuratezza di un test diagnostico ad interpretazione oggettiva, può essere eliminato il quesito guida relativo al *blinding* di chi interpreta i risultati del test in studio rispetto allo standard di riferimento. Gli autori della RS dovrebbero evitare di aggiungere troppi quesiti guida per non rendere troppo complesso lo strumento. Una volta definito il contenuto, è necessario definire precise istruzioni per il rating della RS. Lo strumento dovrebbe essere testato in maniera indipendente da almeno due persone; se la concordanza è buona, lo strumento può essere utilizzato per valutare tutti gli studi inclusi, altrimenti può rendersi necessario un ulteriore affinamento.

FASE 3. DIAGRAMMA DI FLUSSO

Lo step successivo consiste nel rivedere il diagramma di flusso dello studio primario o costruirlo se non è riportato o inadeguato. Il diagramma di flusso faciliterà la valutazione del rischio di bias e dovrebbe fornire informazioni sul metodo di reclutamento dei pazienti (es. serie consecutiva di pazienti con sintomi specifici che fanno sospettare la condizione di interesse, oppure su casi e controlli), la sequenza di esecuzione del test e il numero di pazienti sottoposti al test in studio e allo standard di riferimento. Il diagramma di flusso può essere disegnato a mano perché questo step non deve essere riportato nella valutazione con QUADAS-2. La figura 2 riporta l'esempio di un diagramma di flusso relativo a uno studio primario che ha valutato l'accuratezza diagnostica del peptide natriuretico di tipo B per la diagnosi di scompenso cardiaco.

FASE 4. VALUTAZIONE DEL RISCHIO DI BIAS E DELL'APPLICABILITÀ

Rischio di bias. La prima parte di ogni dominio riguarda i bias e comprende tre sezioni: informazioni utilizzate per valutare il rischio di bias, quesiti guida e valutazione del rischio di bias. Registrando le informazioni utilizzate per la valutazione ("supporto alla valutazione"), si è cercato di renderla trasparente e di facilitare la discussione tra i revisori che effettuano le valutazioni in maniera indipendente[5]. I quesiti guida aggiuntivi aiutano a formulare un giudizio: la risposta può essere "sì", "no", "non chiaro", tenendo conto che "sì" corrisponde a un basso rischio di bias.

Il rischio di bias viene giudicato come "basso", "elevato" o "non chiaro". Se a tutti i quesiti guida relativi ad un dominio è stato risposto "sì", il rischio di bias può essere ritenuto "basso". Se ad ogni quesito guida è stato risposto "no" questo evidenzia un rischio potenziale di bias. I revisori, quindi, devono utilizzare le istruzioni sviluppate nella fase 2 per stimare il rischio di bias. La categoria "non chiaro" dovrebbe essere usata solo quando i dati riportati sono insufficienti per consentire una valutazione.

Applicabilità. Le sezioni relative all'applicabilità sono strutturate in modo analogo a quelle dei bias, ma non includono i quesiti guida. I revisori devono registrare le informazioni su cui viene basato il giudizio di applicabilità ed esprimere le loro perplessità quando il disegno dello studio non corrisponde al quesito della RS. Il rischio di problemi di applicabilità è classificato come "basso", "elevato", "non chiaro". Le valutazioni sull'applicabilità dovrebbero sempre fare riferimento alla prima fase che riporta il quesito della RS. Anche in questo caso la categoria "non chiaro" dovrebbe essere utilizzata quando i dati riportati sono insufficienti. Le sezioni successive forniscono per ciascun dominio del QUADAS-2 una spiegazione sintetica dei quesiti guida per valutare il rischio di bias e i problemi di applicabilità.

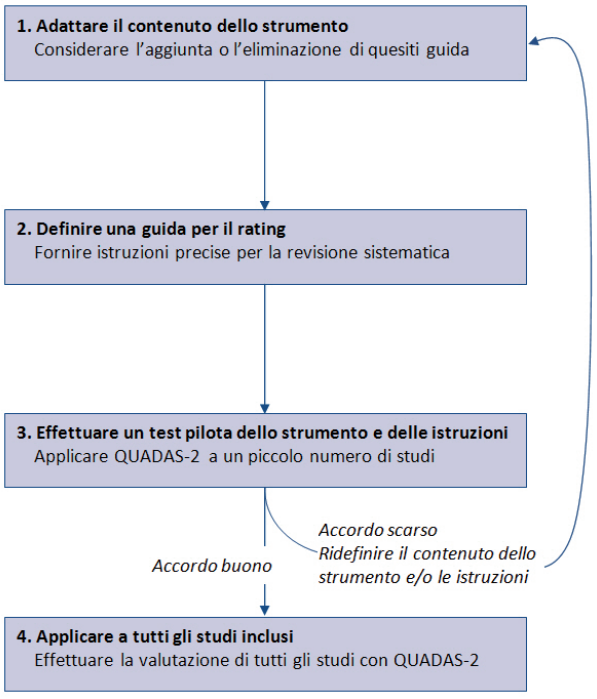

Figura 1. Come adattare lo strumento QUADAS-2 alla revisione sistematica

evidence
open access journal published by the GIMBE Foundation
Standards & Guidelines
OPEN ACCESS

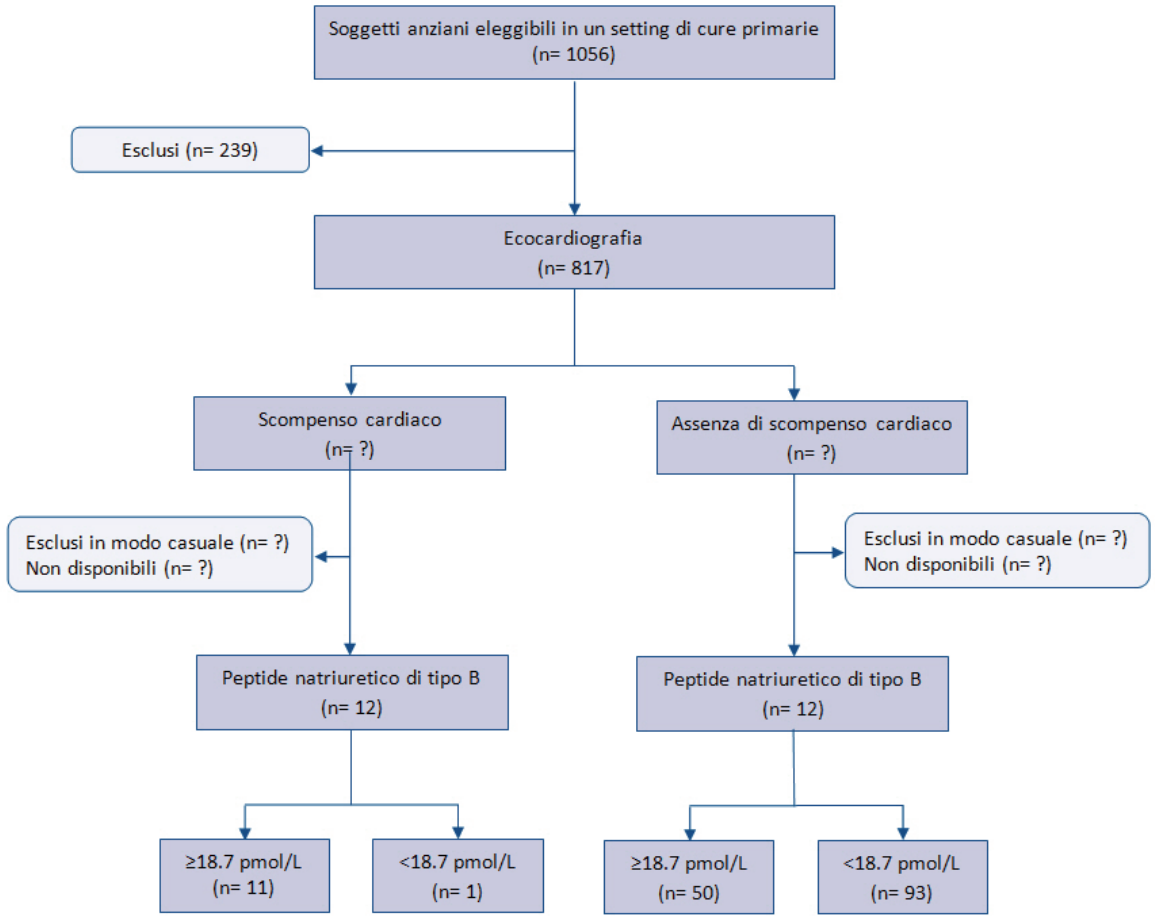

Figura 2. Diagramma di flusso basato su uno studio diagnostico di coorte che ha valutato l'accuratezza del peptide natriuretico di tipo B in per la diagnosi di scompenso cardiaco. Smith H, Pickering RM, Struthers A, Simpson I, Mant D. Biochemical diagnosis of ventricular dysfunction in elderly patients in general practice: observational study. BMJ 2000;320:906-8.

Dominio 1. Selezione dei pazienti

Rischio di bias: la selezione dei pazienti potrebbe essere fonte di bias?

Quesito guida 1. È stato arruolato un campione consecutivo o casuale di pazienti?

Quesito guida 2. È stato evitato un disegno di studio caso-controllo?

Quesito guida 3. Lo studio ha evitato esclusioni di pazienti inappropriate?

Uno studio di accuratezza diagnostica dovrebbe idealmente arruolare un campione consecutivo, o casuale, di pazienti eleggibili con il sospetto di malattia, per evitare potenziali bias. Gli studi che effettuano esclusioni inappropriate, ad esempio escludendo pazienti "difficili da diagnosticare" possono determinare stime di accuratezza diagnostica troppo ottimistiche. Ad esempio, in una RS sugli anticorpi anti-CCP per la diagnosi di artrite reumatoide, alcuni studi hanno arruolato pazienti consecutivi con diagnosi confermate. Questi studi hanno mostrato una maggiore sensibilità del test anti-CCP rispetto ad altri che includevano sia pazienti con sospetta malattia, ma nei quali la diagnosi non era stata confermata, sia quelli "difficili da diagnosticare"[10]. Analogamente, gli studi che arruolano pazienti con malattia nota e un gruppo di controllo senza la condizione target possono sovrastimare l'accuratezza diagnostica[7,11]. L'esclusione di pazienti con segnali di allarme (*red flags*) per la condizione target, che possono essere più facili da diagnosticare, può invece sottostimare l'accuratezza diagnostica.

Applicabilità: pazienti inclusi e setting di arruolamento potrebbero non corrispondere al quesito della RS?

Problemi di applicabilità si possono verificare se i pazienti inclusi nello studio sono diversi rispetto a quelli previsti dal quesito della RS, in termini di gravità della condizione target, caratteristiche demografiche, presenza di diagnosi differenziali o comorbidità, setting dello studio e precedenti protocolli di test. Ad esempio, tumori di maggiori dimensioni, rispetto a quelli più piccoli, sono più facilmente identificabili con i test di imaging e infarti del miocardio più estesi, rispetto a quelli più circoscritti, de-

terminano livelli più elevati di enzimi cardiaci rispetto. Di conseguenza una loro più frequente identificazione aumenta la sensibilità dei test[3].

Dominio 2. Test in studio

Rischio di bias: l'esecuzione o l'interpretazione del test in studio potrebbe essere fonte di bias?

Quesito guida 1. I risultati del test in studio sono stati interpretati senza conoscere i risultati dello standard di riferimento?

Questo item è simile alla definizione di *blinding* negli studi sperimentali: infatti, l'interpretazione del test in studio può essere influenzata dalla conoscenza dei risultati dello standard di riferimento[7]. Il potenziale bias è causato dalla soggettività interpretativa del test in studio e dalla sequenza temporale. Se il test in studio è sempre eseguito e interpretato prima dello standard di riferimento, la risposta al quesito è affermativa.

Quesito guida 2. Il valore soglia eventualmente utilizzato è stato predefinito?

La selezione di un valore soglia del test per ottimizzare sensibilità e/o specificità può determinare performance del test eccessivamente ottimistiche, che rischiano di essere più deboli in un campione indipendente di pazienti in cui viene utilizzato lo stesso valore soglia[12].

Applicabilità: il test in studio, la sua esecuzione o interpretazione potrebbero differire dal quesito della RS?

Le differenze relative alla tecnologia diagnostica, la sua esecuzione o interpretazione possono influenzarne le stime di accuratezza diagnostica. Se i metodi del test in studio variano rispetto a quelli definiti dal quesito della RS, possono esserci problemi di applicabilità. Ad esempio, un trasduttore a ultrasuoni ad elevata frequenza aumenta la sensibilità per valutare i pazienti con trauma addominale[13].

Dominio 3. Standard di riferimento

Rischio di bias: lo standard di riferimento, la sua conduzione o interpretazione potrebbero essere fonte di bias?

Quesito guida 1. Lo standard di riferimento è adeguato per classificare correttamente la condizione target?

Considerato che le stime di accuratezza dei test si basano sul presupposto che lo standard di riferimento abbia sensibilità e specificità del 100%, si assume che le differenze rilevate tra risultati del test in studio e quelli dello standard di riferimento siano conseguenti a una minore accuratezza diagnostica del test in studio[14,15].

Quesito guida 2. I risultati dello standard di riferimento sono stati interpretati senza conoscere i risultati del test in studio?

Il quesito è simile a quello relativo all'interpretazione del test in studio: il potenziale bias può originare dal conoscere il risultato dello standard di riferimento[7].

Applicabilità: la condizione target definita dallo standard di riferimento potrebbe non corrispondere al quesito della RS?

Lo standard di riferimento può non essere affetto da bias, ma la condizione target che identifica può differire da quella prevista dal quesito della RS. Ad esempio, quando si definisce l'infezione delle vie urinarie lo standard di riferimento è generalmente basato sull'urinocoltura, ma il valore soglia di positività può variare[16].

Dominio 4. Flusso e timing

Rischio di bias. Il flusso di pazienti potrebbe essere fonte bias?

Quesito guida 1. L'intervallo temporale tra l'esecuzione del test in studio e dello standard di riferimento è adeguato?

Idealmente i risultati del test in studio e dello standard di riferimento dovrebbero essere ottenuti sugli stessi pazienti nello stesso momento. In caso di ritardo, oppure se tra l'esecuzione del test in studio e lo standard di riferimento è stato iniziato un trattamento, può verificarsi un'errata classificazione dei pazienti conseguente al miglioramento/peggioramento della condizione target. La durata dell'intervallo può determinare un elevato rischio di bias che varia in relazione alla condizione target: ad esempio un ritardo di un paio di giorni può non essere un problema per le malattie croniche, mentre per le malattie infettive acute anche un breve ritardo può essere rilevante. Viceversa, se lo standard di riferimento richiede un periodo di follow-up, questo deve essere sufficientemente lungo per valutare la presenza/assenza della condizione target. Ad esempio, per valutare l'accuratezza diagnostica della risonanza magnetica nella diagnosi precoce della sclerosi multipla, è necessario un follow-up di almeno 10 anni per essere certi che tutti i pazienti soddisfino i criteri diagnostici per la sclerosi multipla[17].

Quesito guida 2. Tutti i pazienti hanno ricevuto lo stesso standard di riferimento?

Il bias di verifica (*verification bias*) avviene quando non tutti i pazienti hanno una conferma diagnostica con lo stesso standard di riferimento. Se i risultati del test in studio condizionano la decisione sullo standard di riferimento da utilizzare o utilizzato, l'accuratezza diagnostica può essere soggetta a bias[11,18]. Ad esempio, uno studio che ha valutato l'accuratezza del D-dimero per la diagnosi di embolia polmonare, ha effettuato la scintigrafia ventilatoria-perfusoria (standard di riferimento 1) nei soggetti positivi al test e utilizzato solo un follow-up clinico nei soggetti con test negativo (standard di riferimento 2). Questo può comportare l'errata classificazione di alcuni falsi negativi come veri negativi: pazienti con embolia polmonare negativi al test in studio, se persi al follow-up, possono essere classificati come soggetti senza embolia polmonare, determinando una sovrastima di sensibilità e specificità.

evidence
open access journal published by the GIMBE Foundation
Standards & Guidelines
OPEN ACCESS

Tabella 2. Esempio di presentazione tabellare per riportare la valutazione con QUADAS-2

Studio	Rischio di bias				Problemi di applicabilità		
	Selezione dei pazienti	Test in studio	Standard di riferimento	Flusso e timing	Selezione dei pazienti	Test in studio	Standard di riferimento
1	☺	☺	☺	☺	☹	☺	☺
2	☺	☺	☺	☺	☹	☺	☺
3	☹	☹	☺	☺	☹	☺	☺
4	☹	☹	☺	☺	☹	☺	☺
5	☹	?	☺	☺	☹	☺	☺
6	☹	?	☺	☺	☹	?	☺
7	☹	?	☺	☺	☹	☺	☺
8	☹	?	☺	☺	☹	?	☺
9	☹	?	☺	☺	☹	☺	☺
10	☹	?	☺	☹	☹	☺	☺
11	☺	?	☺	☹	☺	☺	☺

☺ Rischio basso ☹ Rischio elevato ? Rischio non chiaro

Quesito guida 3. Tutti i pazienti arruolati sono stati inclusi nell'analisi?

Tutti i pazienti arruolati nello studio dovrebbero essere inclusi nell'analisi[19], visto che la discrepanza tra il numero di pazienti arruolati e quelli inclusi nella tabella 2x2 dei risultati è fonte di potenziale bias. Ad esempio, i pazienti persi al follow-up differiscono sistematicamente da quelli che rimangono nello studio.

INTEGRARE LE VALUTAZIONI DI QUADAS-2 NELLE REVISIONI SISTEMATICHE DEGLI STUDI DI ACCURATEZZA DIAGNOSTICA

QUADAS-2 non dovrebbe essere usato per generare uno "score di qualità" cumulativo, a causa delle criticità ben note con questi score[20,21]. Se ad uno studio viene assegnato lo score come "basso" rispetto a tutte le sezioni relative ai bias o all'applicabilità, è opportuno formulare una valutazione complessiva di "basso rischio di bias" o "limitati problemi di applicabilità". Se uno studio è giudicato a rischio "elevato" o "non chiaro" su una o più sezioni, allora può essere giudicato "a rischio di bias" o "con problemi di applicabilità".

Le RS dovrebbero prevedere almeno una sintesi dei risultati della valutazione con QUADAS-2 per tutti gli studi inclusi, ovvero il numero di studi per i quali il rischio di bias e di problemi di applicabilità per ciascun dominio è stato valutato "basso", "elevato" o "non chiaro". I revisori possono scegliere di evidenziare gli studi in cui hanno dato costantemente una valutazione positiva o negativa rispetto a specifici quesiti guida. La tabella 2 e la figura 3 riportano esempi su come presentare in maniera sintetica le valutazioni con QUADAS-2.

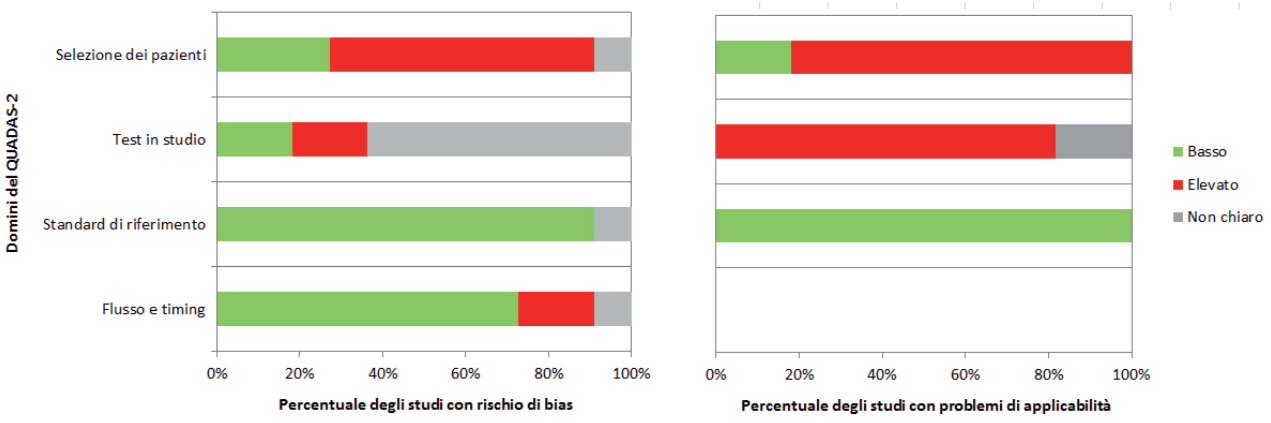

Figura 3. Esempio di presentazione grafica per riportare la valutazione con QUADAS-2

I revisori possono scegliere di restringere l'analisi primaria così da includere solo gli studi a basso rischio di bias e/o con limitati problemi di applicabilità. Può essere appropriato restringere l'inclusione degli studi nella RS in relazione a criteri omogenei, ma spesso è preferibile esaminare tutte le evidenze rilevanti per poi individuare possibili cause di eterogeneità[17,22]. Un'analisi per sottogruppi e/o di sensibilità tra gli studi può essere condotta per documentare la variazione delle stime sull'accuratezza diagnostica del test in studio rispetto a un rischio di bias elevato, basso o non chiaro. I domini o i quesiti guida possono essere inclusi come item in un'analisi di meta-regressione, per valutare la loro associazione rispetto all'accuratezza stimata.

Il sito web QUADAS (www.quadas.org) contiene lo strumento QUADAS-2, informazioni sul training, una banca di quesiti guida aggiuntivi, una guida più dettagliata per ogni sezione, esempi di valutazioni completate con QUADAS-2 e risorse scaricabili, tra cui un database Microsoft Access per l'estrazione dei dati, un foglio Excel per la produzione di presentazioni grafiche dei risultati e modelli di tabelle in Word per la sintesi dei risultati.

DISCUSSIONE

L'attenta valutazione della qualità degli studi inclusi è fondamentale per condurre RS degli studi di accuratezza diagnostica. È stato utilizzato un rigoroso processo *evidence-based* per sviluppare lo strumento QUADAS-2 dal QUADAS, già ampiamente utilizzato. Lo strumento QUADAS-2 offre elementi aggiuntivi e perfezionati, inclusa la distinzione tra bias e applicabilità. Individua 4 domini chiave supportati da quesiti guida per facilitare la valutazione e la classificazione del rischio di bias, i problemi di applicabilità e le modalità di valutazione degli studi che prevedono il follow-up come standard di riferimento.

QUADAS-2 rappresenta una notevole evoluzione dello strumento originale: sarebbe auspicabile estendere QUADAS-2 per consentire la valutazione di studi che confrontano più test in studio, ma le evidenze scientifiche in questo ambito sono ancora insufficienti ed è necessario pianificare ulteriori ricerche. Auspicando che QUADAS-2 aiuti a sviluppare una robusta *evidence-base* per test e procedure diagnostiche, sono benvenuti commenti e feedback tramite il sito web QUADAS.

RINGRAZIAMENTI

Gli autori ringraziano i revisori che hanno partecipato al test pilota: Gianni Virgilli, Vittoria Murro, Karen Steingardt, Laura Flores, Beth Shaw, Toni Tan, Kurinchi Gurusamy, Mario Cruciani, Lee Hooper e Catherine Jameson. Gli autori ringraziano inoltre il Cochrane Collaboration's Diagnostic Test Accuracy Working Group per la collaborazione a parte delle attività di questo progetto.

MEMBRI DEL GRUPPO QUADAS-2

University of Oxford: Doug Altman, Susan Mallett*; Memorial Sloan-Kettering Cancer Center: Colin Begg; University of Bristol: Rebecca Beynon, Jonathan A.C. Sterne*, Penny F. Whiting*; University of Amsterdam: Patrick M.M. Bossuyt*, Mariska M.G. Leeflang*, Jeroen Lijmer; American College of Physicians: John Cornell; University of Birmingham: Clare Davenport, Jonathan J. Deeks*, Khalid Khan; Bond University: Paul Glasziou; University of Sydney: Rita Horvath, Les Irwig, Petra Macaskill; University of Exeter: Chris Hyde; Maastricht University: Jos Kleijnen; University Medical Center Utrecht: Karel G.M. Moons, Johannes B. Reitsma*; Basel Institute of Clinical Epidemiology and Biostatistics: Heike Raatz; University of Bern: Anne W.S. Rutjes*; National Institute for Health and Clinical Excellence: Beth Shaw, Toni Tan; Keele University: Danielle van der Windt; Università di Firenze: Gianni Virgili; Kleijnen Systematic Reviews: Marie E. Westwood*.

*membri del comitato direttivo.

MATERIALE SUPPLEMENTARE

Strumento QUADAS-2
Template per la presentazione grafica
QUADAS-2 Database

CONTRIBUTO DEGLI AUTORI

Ideazione e disegno dello studio: Penny F Whiting, Anne WS Rutjes, Johannes B Reitsma, Mariska MG Leeflang, Jonathan AC Sterne, Patrick MM Bossuyt.

Analisi e interpretazione dei dati: Penny F Whiting, Anne WS Rutjes, Marie E Westwood, Jonathan J Deeks, Johannes B Reitsma, Jonathan AC Sterne, Patrick MM Bossuyt.

Stesura del manoscritto: Penny F Whiting, Marie E Westwood, Susan Mallett, Mariska MG Leeflang, Jonathan AC Sterne.

Revisione critica di importanti contributi intellettuali: Anne WS Rutjes, Marie E Westwood, Jonathan J Deeks, Johannes B Reitsma, Mariska MG Leeflang, Jonathan AC Sterne, Patrick MM Bossuyt.

Approvazione finale del manoscritto: Penny F Whiting, Anne WS Rutjes, Marie E Westwood, Susan Mallett, Jonathan J Deeks, Johannes B Reitsma, Mariska MG Leeflang, Jonathan AC Sterne, Patrick MM Bossuyt.

Consulenza statistica: Jonathan J Deeks, Jonathan AC Sterne.

Fundraising: Penny F Whiting, Jonathan J Deeks, Jonathan AC Sterne.

Supporto amministrativo, tecnico e logistico: Jonathan AC Sterne.

Raccolta e assemblamento dei dati: Penny F Whiting, Anne WS Rutjes, Marie E Westwood, Johannes B Reitsma, Mariska MG Leeflang.

evidence
open access journal published by the GIMBE Foundation

Standards & Guidelines OPEN ACCESS

NOTE ALLA VERSIONE ITALIANA

La Fondazione GIMBE ha realizzato la traduzione italiana dell'articolo senza alcun finanziamento istituzionale o commerciale.

L'*American College of Physicians*, che ha autorizzato la traduzione dell'articolo a fini non commerciali, non si assume alcuna responsabilità per l'accuratezza della traduzione.

TEAM CHE HA REALIZZATO LA VERSIONE ITALIANA

Responsabile scientifico
Antonino Cartabellotta, Fondazione GIMBE

Coordinamento editoriale
Marco Mosti, Fondazione GIMBE

Traduzione
Federica Riccio, Specialista in Igiene e Medicina Preventiva. Casale Monferrato (AL)

Revisione editoriale
Elena Cottafava, Fondazione GIMBE
Roberto Luceri, Fondazione GIMBE

BIBLIOGRAFIA

1. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
2. Willis BH, Quigley M. Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review. BMC Med Res Methodol 2011;11:27.
3. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ, Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2009. Disponibile a: http://dta.cochrane.org/handbook-dta-reviews. Ultimo accesso: 2 febbraio 2016.
4. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med 2010;7:e1000217.
5. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomized trials. BMJ. [Forthcoming]
6. Whiting P, Rutjesw AW, Westwood M, Mallett S, Leefland M, Reitsma JB, et al. Updating QUADAS: evidence to inform the development of QUADAS-2. 2010. Disponibile a: www.bris.ac.uk/quadas/resources/quadas2reportv4.pdf. Ultimo accesso: 2 febbraio 2016.
7. Whiting P, Rutjes AW, Reitsma JB, Glas AS, Bossuyt PM, Kleijnen J. Sources of variation and bias in studies of diagnostic accuracy: a systematic review. Ann Intern Med 2004;140:189-202.
8. Bossuyt PM, Leeflang MM. Chapter 6: Developing criteria for including studies. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration. 2009. Disponibile a: http://dta.cochrane.org/handbook-dta-reviews. Ultimo accesso: 2 febbraio 2016.
9. Leeflang MM, Deeks JJ, Gatsonis C, Bossuyt PM; Cochrane Diagnostic Test Accuracy Working Group. Systematic reviews of diagnostic test accuracy. Ann Intern Med 2008;149:889-97.
10. Whiting PF, Smidt N, Sterne JA, Harbord R, Burton A, Burke M, et al. Systematic review: accuracy of anti-citrullinated peptide antibodies for diagnosing rheumatoid arthritis. Ann Intern Med 2010;152:456-64.
11. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999;282:1061-6.
12. Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 2008;54:729-37.
13. Stengel D, Bauwens K, Rademacher G, Mutze S, Ekkernkamp A. Association between compliance with methodological standards of diagnostic research and reported test accuracy: meta-analysis of focused assessment of US for trauma. Radiology 2005;236:102-11.
14. Biesheuvel C, Irwig L, Bossuyt P. Observed differences in diagnostic test accuracy between patient subgroups: is it real or due to reference standard misclassification? Clin Chem 2007;53:1725-9.
15. van Rijkom HM, Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries—a meta-analysis. Caries Res 1995;29:364-70.
16. Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, et al. Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model. Health Technol Assess 2006;10:iii-iv, xi-xiii, 1-154.
17. Whiting P, Harbord R, Main C, Deeks JJ, Filippini G, Egger M, et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review. BMJ 2006;332:875-84.
18. Rutjes A, Reitsma J, Di NM, Smidt N, Zwinderman A, Van RJ, et al. Bias in estimates of diagnostic accuracy due to shortcomings in design and conduct: empirical evidence [Abstract]. Presented at XI Cochrane Colloquium: Evidence, Health Care and Culture, Barcelona, Spain, 26–31 October 2003. Abstract 45.
19. Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and Presenting Results. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0. The Cochrane Collaboration, 2010. Disponibile a: http://dta.cochrane.org/handbook-dta-reviews. Ultimo accesso: 2 febbraio 2016.
20. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054-60
21. Whiting P, Harbord R, Kleijnen J. No role for quality scores in systematic reviews of diagnostic accuracy studies. BMC Med Res Methodol 2005;5:19.
22. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006;6:9.