Rassegna

L'iniziativa STARD per la produzione di studi completi ed accurati di accuratezza diagnostica*

P.M. Bossuyt[1], J.B. Reitsma[1], D.E. Bruns[2,3], C.A. Gatsonis[4], P.P. Glasziou[5], L.M. Irwig[6], J.G. Lijmer[1], D. Moher[7], D.Rennie[8,9] , H.C.W. De Vet[10] per lo STARD Group

[1] *Department of Clinical Epidemiology and Biostatistics, Academic Medical Center University of Amsterdam, 1100 DE Amsterdam, The Netherlands*
[2] *Department of Pathology, University of Virginia, Charlottesville, VA 22903*
[3] *Clinical Chemistry, Washington, DC 20037*
[4] *Centre for Statistical Sciences, Brown University, Providence, RI 02912*
[5] *Centre for General Practice, University of Queensland, Herston QLD 4006, Australia*
[6] *Department of Public Health & Community Medicine, University of Sydney, Sydney NSW 2006, Australia*
[7] *Chalmers Research Group, Ottowa, Ontario, K1N 6M4 Canada*
[8] *Institute for Health Policy Studies, University of California, San Francisco, San Francisco, CA 9411*
[9] *Journal of the American Medical Association, Chicago, IL 60610*
[10] *Institute for Research in Extramural Medicine, Free University, 1081 BT Amsterdam, The Netherlands*

***Il documento è stato tradotto con il permesso del direttore di Clinical Chemistry David E. Bruns da Romolo M. Dorizzi**

Background: Per comprendere i risultati di studi di accuratezza diagnostica, i lettori devono comprendere come sono progettati, come sono condotti, come sono eseguite le analisi e come sono prodotti i risultati. Questo obiettivo può essere raggiunto solo con la completa trasparenza degli autori

Obiettivo*:* Migliorare l'accuratezza e la completezza della produzione di articoli relativi alla accuratezza per consentire ai lettori di valutare la possibilità di *bias* e di valutare se i risultati dello studio sono generalizzabili.

Metodi: Il comitato direttivo dello Standards for Reporting of Diagnostic Accuracy (Standard per la produzione di articoli relativi alla Accuratezza Diagnostica, STARD) ha condotto una ricerca della letteratura, allo scopo di individuare pubblicazioni dedicate alla conduzione corretta ed alla pubblicazione di studi sugli esami diagnostici, di ricavarne i punti salienti e di raccoglierli in una lista esaustiva. Ricercatori, direttori di giornali e membri di organizzazioni professionali hanno ridotto questa lista nel corso di una riunione di consenso di due giorni convocata allo scopo di produrre una lista di controllo e un diagramma di flusso generale dedicato agli studi di accuratezza diagnostica.

Risultati: La ricerca della letteratura per linee guida dedicate alla diagnostica ha individuato 33 liste di controllo da cui è stata estratta una lista di 75 punti potenziali. La riunione di consenso ha ridotto la lista a 25 punti usando, ogni volta che era disponibile, evidenza sul bias. Un prototipo di diagramma di flusso fornisce informazioni sulle modalità di reclutamento, l'ordine di esecuzione degli esami e il numero di pazienti sottoposti all'esame in valutazione, allo standard di riferimento o ad entrambi.

Conclusioni: La valutazione di una ricerca è condizionata dalla esposizione chiara ed accurata. Se i giornali medici adottano la lista di controllo, la qualità degli articoli dedicati alla accuratezza diagnostica migliorerà a vantaggio di clinici, ricercatori, revisori, giornali e pazienti.

Il mondo degli esame diagnostici è in continua evoluzione. Nuovi esami sono sviluppati a velocità molto rapida e la tecnologia di quelli in uso è continuamente perfezionata. Risultati esagerati o con bias ottenuti con studi diagnostici progettati male e descritti in articoli di cattiva qualità possono portare alla loro diffusione prematura e portare i clinici ad iniziative terapeutiche errate. Una valutazione rigorosa degli esami diagnostici prima della loro introduzione nella pratica clinica potrebbe non solo ridurre le conseguenze cliniche dannose ma anche ridurre i costi sanitari, evitando esami non necessari.

Gli studi per definire l'accuratezza diagnostica di un esame costituiscono una parte vitale di tale processo di valutazione[1-3].

Negli studi di accuratezza diagnostica, i risultati dell'esame da valutare sono confrontati con quelli ottenuti con un metodo di riferimento (i due esami sono eseguiti in soggetti in cui esiste il sospetto che siano affetti dalla condizione di interesse). Il termine *esame* si riferisce a qualsiasi metodo per ottenere informazioni aggiuntive sullo stato di salute di un soggetto. Comprende cioè informazioni provenienti dall'anamnesi, dalla visita, da esami di laboratorio, da esami di diagnostica per immagini, esami funzionali ed anatomo patologici. La condizione che interessa, la condizione bersaglio (*target)* può riferirsi ad una malattia particolare o ad una qualsiasi altra condizione identificabile che possa portare ad un intervento clinico, come l'esecuzione di un ulteriore esame diagnostico o l'inizio, la modifica o l'interruzione di una terapia. In questo ambito, lo *standard di riferimento* è considerato il miglior metodo disponibile per stabilire la presenza o l'assenza della condizione di interesse. La presenza della condizione target può essere stabilita da un singolo metodo o da una combinazione di metodi. Può comprendere esami di laboratorio, di diagnostica per immagini, anatomo patolologici, procedure di monitoraggio dei soggetti nel tempo (*follow-up*). Il termine *accuratezza* si riferisce al grado di concordanza (espressa quantitativamente) tra l'informazione ottenuta dall'esame in valutazione e lo standard di riferimento. L'accuratezza diagnostica può essere espressa in molti modi: sensibilità e specificità, quoziente di probabilità (*likelihood ratio), odds ratio* diagnostici o area sottostante una curva ROC[4-6].

Esistono numerose minacce potenziali alla validità interna ed esterna di uno studio di accuratezza diagnostica. Un esame di questo tipo di studi - pubblicati tra il 1978 e il 1993 in quattro dei più importanti giornali medici - ha rivelato che la qualità metodologica era, nel migliore dei casi, mediocre[7]. Le valutazioni erano comunque ostacolate dal fatto che molti articoli mancavano di informazioni su elementi chiave del disegno, dell'esecuzione e dell'interpretazione degli studi diagnostici[7]. Gli autori di alcune metanalisi hanno confermato l'assenza di informazioni critiche circa il disegno e l'esecuzione degli studi[8,9]. Come per qualsiasi altro tipo di ricerca, gli errori nel disegno dello studio possono portare a risultati affetti da *bias*. Un articolo ha dimostrato che studi diagnostici con certe caratteristiche d'impostazione sono associati a stime dell'accuratezza diagnostica affette da bias ed ottimistiche, rispetto a studi senza tali difetti[10].

Al *Cochrane Colloquium* di Roma del 1999, il *Cochrane Diagnostic and Screening Test Methods Working Group* ha discusso la bassa qualità metodologica degli studi diagnostici e la bassa qualità degli articoli relativi. Secondo il *Working Group* il primo passo per correggere questi problemi era migliorare la qualità di questi articoli. Dopo il successo dell'iniziativa CONSORT[11-13], il *Working Group* si è concentrato sulla produzione di una lista di controllo di punti, da inserire in un articolo dedicato alla accuratezza diagnostica.

L'obiettivo dell'iniziativa *Standards for Reporting Diagnostic Accuracy* (STARD) è quella di migliorare la qualità degli articoli dedicati alla accuratezza diagnostica. Un articolo completo ed accurato permette al lettore di individuare i potenziali *bias* nello studio (**validità interna**) e di valutare la possibilità di generalizzare e applicare i risultati (**validità esterna**).

Materiale e metodi

Il comitato direttivo dello STARD (*vedi appendice con il nome dei membri e altri dettagli*) ha condotto preliminarmente una ampia ricerca per identificare gli articoli dedicati alla conduzione ed alla preparazione di articoli dedicati agli studi diagnostici. Sono stati utilizzati Medline, Embase, BIOSIS e la banca dati metodologica Cochrane Collaboration fino al luglio 2000. Inoltre, i membri del comitato direttivo hanno esaminato la bibliografia degli articoli individuati, i loro archivi personali e hanno contattato altri esperti nel campo. Dopo un esame di tutti gli articoli rilevanti hanno ricavato una ampia lista di potenziali punti della lista di controllo.

E' stata poi organizzata una riunione di consenso di due giorni, alla quale sono stati invitati esperti provenienti dalla ricerca, dall'editoria, che avevano condotto studi metodologici oltre a membri di organizzazioni professionali. L'obiettivo della conferenza era di ridurre l'elenco dei punti potenziali, dove opportuno, e di discutere sul formato e sulla organizzazione della lista di controllo.

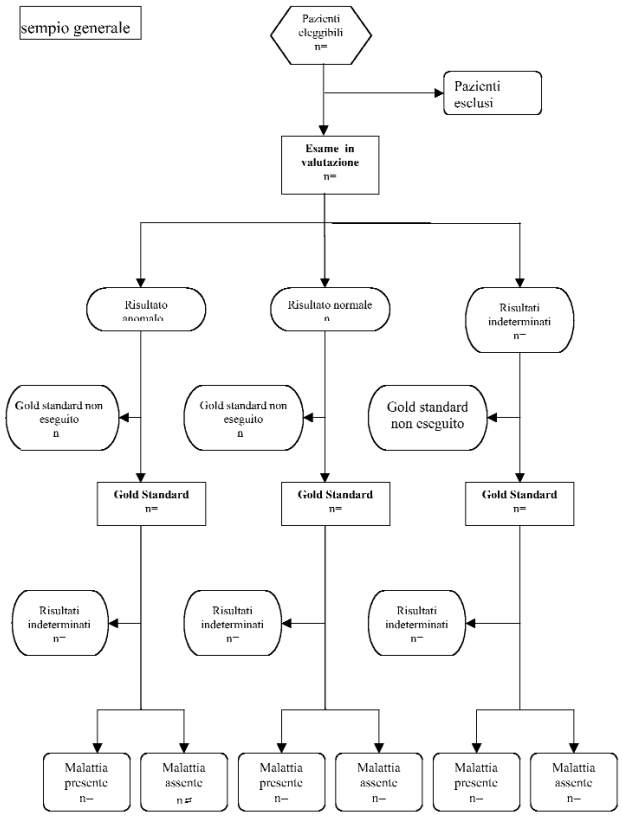
La conferenza ha compreso sessioni ristrette e sessioni plenarie. I partecipanti ad ogni sessione ristretta si sono concentrati sull'esame di un gruppo di punti affini della lista; i loro suggerimenti sono stati poi discussi nelle sessioni plenarie. Alla fine della prima giornata, sulla base dei suggerimenti dei piccoli gruppi e dei suggerimenti delle sessioni plenarie è stata preparata una prima bozza della lista di controllo della STARD. Alla fine del primo giorno era già disponibile una prima stesura della *lista di controllo*. Tutti hanno partecipato alla riunione del giorno successivo in cui sono state proposte modifiche. I membri del gruppo STARD potevano proporre ulteriori modifiche attraverso posta elettronica. La lista di controllo e il diagramma di flusso prodotti nel corso della conferenza espressa dalla conferenza sono state poi valutate sul campo da utilizzatori potenziali e sono stati poi raccolti ulteriori commenti. Questa versione è stata resa disponibile sul sito web CONSORT richiedendo altri commenti. Il comitato

direttivo STARD ha discusso tutti i commenti e ha assemblato la lista di controllo finale.

Risultati

La ricerca delle linee guida per gli studi diagnostici presenti in letteratura ha individuato 33 liste di controllo. Sulla base di questo materiale e dei suggerimenti del comitato direttivo e dei membri del gruppo STARD è stata ottenuta una lista di 75 punti. Nella riunione di consenso del 16 e 17 settembre 2000, il gruppo di lavoro ha ridotto il numero dei punti a 25 e ha introdotto importanti modifiche nel testo e nella struttura.
Il gruppo STARD ha ricevuto importanti commenti ed osservazioni durante le fasi successive alla conferenza e ha prodotto la versione della lista di controllo STARD presentata nella tabella I.
Il diagramma di flusso dà informazioni sui metodi di reclutamento dei pazienti (ad esempio serie consecutiva di soggetti con sintomi specifici, oppure con modalità caso-controllo), sull'ordine di esecuzione delle determinazioni, sul numero di pazienti sottoposti all'esame in valutazione (esame indice) e all'esame di riferimento (vedi Figura 1). Quello mostrato è un prototipo di diagramma di flusso relativo al disegno più comunemente impiegato nella ricerca diagnostica. Esempi che riflettono altri disegni possono essere trovati sul sito web STARD (Vedi www.consort-statement.org.htm)

Figura 1. Prototipo del diagramma di flusso per uno studio sulla accuratezza diagnostica.



Discussione

Lo scopo dell'iniziativa STARD è quella di migliorare la qualità degli articoli dedicati all'accuratezza diagnostica. I punti della lista di controllo e nel diagramma di flusso possono aiutare gli autori nel descrivere i punti essenziali del disegno di uno studio, la sua esecuzione e i suoi risultati.
Abbiamo diviso i punti della lista di controllo nelle **sezioni** classiche **di un articolo** di ricerca medica (Introduzione, Metodi, Risultati, Discussione) anche se l'ordine in cui appaiono nell'articolo può variare.
Il criterio guida nello sviluppare la *lista di controllo* STARD è stato quello di scegliere punti in grado di aiutare i lettori ad individuare il potenziale di *bias* dello studio e valutarne l'applicabilità dei risultati.
Altre due considerazioni generali ne hanno modellato forma e contenuto. Primo: il gruppo STARD ha ritenuto che una *lista di controllo* unica per gli studi di accuratezza in tutti i campi della diagnostica fosse più facilmente diffusa e forse accettata da autori, revisori e direttori di giornali piuttosto che *liste di controllo* diverse per ciascuno di essi. Infatti, benchè la valutazione di un esame di *imaging* differisca da quella di un esame di laboratorio, riteniamo che le differenze siano più di quantità che di qualità. Secondo: questa è una *lista di controllo* specifica, mirata agli studi di accuratezza diagnostica. Non abbiamo inserito aspetti generali sul modo di preparare articoli di ricerca, analoghi alle raccomandazioni contenute negli *Uniform Requirements for Manuscripts submitted to Biomedical Journals*[14].
Tutte le volte che è risultato possibile il gruppo STARD ha basato la decisione di inserire un punto sulla base delle prove che legano il punto a stime non corrette (validità interna) o a variazioni nella misura dell'accuratezza diagnostica (validità esterna). La solidità delle prove va dagli articoli di tipo narrativo, che affrontano principi teorici a quelli che presentano risultati ottenuti con modelli statistici a prove empiriche ricavate da studi diagnostici. Per molti punti le prove sono piuttosto limitate.
Un documento di spiegazione separato spiega il significato e la logica di ogni elemento e riassume brevemente il tipo e la numerosità delle prove[15]. Tale documento favorirà l'uso, la comprensione e la diffusione della *lista di controllo* STARD.
Il gruppo STARD si è impegnato molto nello sviluppo di un diagramma di flusso per studi diagnostici. Un diagramma di flusso è in grado di comunicare informazioni vitali sull'architettura di uno studio e sul flusso dei partecipanti in modo trasparente[16]. Un diagramma di flusso simile è diventato un elemento essenziale degli standard CONSORT per la preparazione di articoli dedicati ai trial randomizzati. Tale diagramma potrebbe essere ancor più essenziale negli studi diagnostici, data la varietà dei disegni impiegati in questo ambito di ricerca. L'uso dei diagrammi di flusso negli articoli che si occupano di

Tabella I. Lista di controllo STARD per gli articoli dedicati all'accuratezza diagnostica

Sezione e argomento	Punto N	Descrizione	Pagina N
TITOLO RIASSUNTO PAROLE CHIAVE	**1**	Identificare l'articolo come studio dell'accuratezza diagnostica (raccomandato il MeSH heading "sensitivity and specificità")	
INTRODUZIONE	**2**	Definire quesiti o obiettivi dello studio come stima dell'accuratezza diagnostica o confronto dell'accuratezza tra esami o gruppi di partecipanti	
METODI		Descrivere	
Partecipanti	**3**	Popolazione dello studio: criteri d'inclusione ed esclusione, sede in cui sono stati raccolti i dati	
	4	Reclutamento dei partecipanti: è basato sulla sintomatologia alla presentazione, sui risultati di esami precedenti o sul fatto che i partecipanti erano stati sottoposti all'esame in studio o a quello di riferimento?	
	5	Campionamento dei partecipanti: è avvenuto su una serie consecutiva di pazienti definiti dai criteri di selezione ai punti 3 e 4? Se no, specificare come sono stati selezionati i pazienti.	
	6	Raccolta dei dati: la raccolta dei dati è stata pianificata prima che l'esame in studio e il riferimento fossero eseguiti (studio prospettico) o dopo (studio retrospettivo)?	
Metodi	**7**	Standard di riferimento e sua logica	
	8	Specifiche tecniche dei materiali e dei metodi impiegati compresi come e quando sono state eseguite le misure (e/o indicazione dei riferimenti per l'esame in studio e il gold standard)	
	9	Definizione e razionale delle unità, dei valori di cut-off e/o delle categorie di risultati dell'esame in studio e di quello di riferimento	
	10	Numero, addestramento ed esperienza delle persone che eseguono ed interpretano l'esame in studio e quello di riferimento	
	11	Chi interpretava l'esame in studio e il gold standard era all'oscuro dei risultati dell'altro esame (in cieco)? erano rese disponibile ai lettori tutte le informazioni cliniche?	
Metodi statistici	**12**	Metodi per calcolare o confrontare i parametri di accuratezza diagnostica; metodi statistici impiegati per quantificare l'incertezza (ad esempio intervalli di confidenza al 95%)	
	13	Metodi per calcolare la riproducibilità, se impiegati	
RISULTATI		Articolo	
Partecipanti	**14**	Data in cui lo studio è stato condotto comprese quelle di inizio e di fine del reclutamento	
	15	Caratteristiche cliniche e demografiche (es. età, sesso, spettro dei sintomi, patologia intercorrente, trattamenti in corso, centri di reclutamento)	
	16	Numero di partecipanti che soddisfano i criteri d'inclusione che non sono stati sottoposti all'esame in valutazione e/o all'esame di riferimento; indicare perché alcuni partecipanti sono stati eventualmente esclusi dall'uno o dall'altro (l'uso di un diagramma di flusso è molto raccomandato)	
Risultati	**17**	Intervallo tra esecuzione dell'esame in valutazione e dell'esame di riferimento; notizie su eventuali trattamenti somministrati nel frattempo	
	18	Spettro di gravità della malattia indagata (definirne i criteri) in coloro che ne sono affetti; descrivere le altre diagnosi nei partecipanti che non ne sono affetti	
	19	Tabella di confronto dei risultati dell'esame in studio e di quelli dell'esame di riferimento (compresi i risultati indeterminati e negativi); per risultati continui distribuzione dei risultati dell'esame in valutazione rispetto a quelli dell'esame di riferimento	
	20	Ogni evento avverso causato dalla esecuzione dell'esame in valutazione o dell'esame di riferimento	
	21	Stime dell'accuratezza diagnostica e misure dell'incertezza statistica (es. intervalli di confidenza al 95%)	
Stime	**22**	Gestione dei risultati indeterminati, risposte mancanti e aberranti dell'esame in valutazione	
	23	Stime della variabilità dell'accuratezza diagnostica tra sottogruppi di partecipanti, lettori o centri (se effettuate)	
	24	Stime della riproducibilità dell'esame (se effettuate)	
DISCUSSIONE	**25**	Discutere l'applicabilità clinica dei risultati dello studio	

studi sull'accuratezza diagnostica comprendono il processo di campionamento e di selezione dei partecipanti (validità esterna), il flusso dei partecipanti in relazione alla temporizzazione, agli outcome degli esami, al numero di soggetti che non sono sottoposti all' esame in studio e/o all'esame di riferimento (potenziale *bias* di *verifica*[17,19], il numero di pazienti in ciascuno stadio dello studio, che fornisce il denominatore corretto per il calcolo delle proporzioni (consistenza interna).

Il gruppo STARD ha il progetto di valutare l'impatto del documento sulla qualità degli articoli pubblicati dedicati all'accuratezza diagnostica degli articoli con uno studio prima-dopo[13]. La *lista di controllo* sarà aggiornata e revisionata regolarmente in base alle prove sulle cause di bias o di variabilità che si rendessero disponibili. Qualsiasi commento di contenuto e di forma è quindi auspicata per migliorare la versione corrente.

Il supporto finanziario per riunire il gruppo STARD è stato fornito in parte da Dutch Health Care Insurance Board, International Federation of Clinical Chemistry, Medical Research Council's Health Services Research Collaboration, ed Academic Medical Center in Amsterdam. Questa iniziativa per migliorare gli articoli dedicati all'accuratezza diagnostica è stata sostenuta dalle molte persone in tutto il modo che hanno inviato commenti alle diverse versioni.

Membri del Comitato Direttivo STARD

Patrick Bossuyt
Academic Medical Center, Dept. of Clinical Epidemiology, Amsterdam, The Netherlands
Constantine Gatsonis
Brown University, Centre for Statistical Sciences Providence, United States of America
Les Irwig
University of Sydney, Dept. of Public Health & Community Medicine, Sydney, Australia
David Moher
Chalmers Research Group, Ottowa, Ontario, Canada
Riekie de Vet
Free University, Institute for Research in Extramural Medicine, Amsterdam, The Netherlands
David Bruns
Clinical Chemistry, Charlottesville, United States of America
Paul Glasziou
Mayne Medical School, Dept. of Social & Preventive Medicine, Herston, Australia
Jeroen Lijmer
Academic Medical Center, Dept. of Clinical Epidemiology, Amsterdam, The Netherlands
Drummond Rennie
Journal of the American Medical Association, Chicago, United States of America

Membri del gruppo STARD

Doug Altman, Institute of Health Sciences, Centre for Statistics in Medicine (Oxford, United Kingdom); Stuart Barton, *British Medical Journal*, BMA House (London,
United Kingdom); Colin Begg, Memorial Sloan-Kettering Cancer Center, Department Epidemiology & Biostatistics (New York, NY); William Black, Dartmouth Hitchcock Medical Center, Department of Radiology (Lebanon, NH);
Harry Bu¨ ller, Academic Medical Center, Department of Vascular Medicine (Amsterdam, The Netherlands); Gregory Campbell, US FDA, Center for Devices and Radiological Health (Rockville, MD); Frank Davidoff, *Annals of Internal Medicine* (Philadelphia, PA); Jon Deeks, Institute of Health Sciences, Centre for Statistics in Medicine (Old Road, United Kingdom); Paul Dieppe, Department Social Medicine, University of Bristol (Bristol, United Kingdom); Kenneth Fleming, John Radcliffe Hospital, (Oxford, United Kingdom); Rijk van Ginkel, Academic Medical Center, Department of Clinical Epidemiology (Amsterdam, The Netherlands); Afina Glas, Academic Medical Center, Department of Clinical Epidemiology (Amsterdam,
The Netherlands); Gordon Guyatt, McMaster University, Clinical Epidemiology and Biostatistics (Hamilton, Canada); James Hanley, McGill University, Department Epidemiology & Biostatistics (Montreal, Canada); Richard Horton, *The Lancet*, (London, United Kingdom); Myriam Hunink, Erasmus Medical Center, Department
of Epidemiology & Biostatistics (Rotterdam, The Netherlands); Jos Kleijnen, NHS Centre for Reviews and Dissemination (York, United Kingdom); Andre Knottnerus,
Maastricht University, Netherlands School of Primary Care Research (Maastricht, The Netherlands); Erik Magid, Amager Hospital, Department of Clinical Biochemistry (Copenhagen, Denmark); Barbara McNeil, Harvard Medical School, Department of Health Care Policy Boston, MA); Matthew McQueen, Hamilton Civic Hospitals, Department of Laboratory Medicine (Hamilton, Canada); Andrew Onderdonk, Channing Laboratory Boston, MA); John Overbeke, *Nederlands Tijdschrift voor Geneeskunde* (Amsterdam, The Netherlands); Christopher
Price, St Bartholomew's - Royal London School of Medicine and Dentistry (London, United Kingdom); Anthony Proto, *Radiology* Editorial Office (Richmond, VA);Hans
Reitsma, Academic Medical Center, Department of Clinical Epidemiology (Amsterdam, The Netherlands); David Sackett, Trout Centre (Ontario, Canada); Gerard Sanders, Academic Medical Center, Department of Clinical Chemistry

(Amsterdam, The Netherlands); Harold Sox, *Annals Internal Medicine* (Philadelphia, PA); Sharon Straus, Mt. Sinai Hospital (Toronto, Canada); Stephan Walter, Mc-
Master University, Clinical Epidemiology and Biostatistics (Hamilton, Canada).

Bibliografia

1. Guyatt GH, Tugwell PX, Feeny DH, Haynes RB, Drummond M. A framework for clinical evaluation of diagnostic technologies. Can Med Assoc J 1986;134:587–94.
2. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991;11:88–94.
3. Kent DL, Larson EB. Disease, level of impact, and quality of research methods. Three dimensions of clinical efficacy assessment applied to magnetic resonance imaging. Invest Radiol 1992; 27:245–54.
4. Griner PF, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. Ann Intern Med 1981;94:557–92.
5. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. The selection of diagnostic tests. In: Sackett D, editor. Clinical epidemiology, 2nd ed. Boston/Toronto/London: Little, Brown and Company; 1991: 47–57.
6. Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8:283–98.
7. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. JAMA 1995;274:645–51.
8. Nelemans PJ, Leiner T, de Vet HCW, van Engelshoven JMA. Peripheral arterial disease: Meta-analysis of the diagnostic performance of MR angiography. Radiology 2000;217:105–14.
9. Devries SO, Hunink MGM, Polak JF. Summary receiver operating characteristic curves as a technique for meta-analysis of the diagnostic performance of duplex ultrasonography in peripheral arterial disease. Acad Radiol 1996;3:361–9.
10. 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.
11. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637–9.
12. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. JAMA 2001;285:1987–91.
13. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials. A comparative before-and afterevaluation. JAMA 2001;285:1992–5.
14. International Committee of Medical Journal Editors. Uniform Requirements for manuscripts submitted to biomedical journals. JAMA. 1997;277:927–34. Also available at: ACP Online, http://www.acponline.org.
15. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD Statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem 2003;49:7–18.
16. Egger M, Ju¨ni P, Barlett C. Value of flow diagrams in reports of randomized controlled trials. JAMA 2001;285:1996–9.
17. Knottnerus JA. The effects of disease verification and referral on the relationship between symptoms and diseases. Med Decis Making 1987;7:139–48.
18. Panzer RJ, Suchman AL, Griner PF. Workup bias in prediction research. Med Decis Making 1987;7:115–9.
19. Begg CB. Biases in the assessment of diagnostic tests. Stat Med 1987;6:411–23.