Methods Research Report

Comprehensive Overview of Methods and Reporting of Meta-Analyses of Test Accuracy

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Methods Research Report

Comprehensive Overview of Methods and Reporting of Meta-Analyses of Test Accuracy

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10055-I

Prepared by:
Tufts Evidence-based Practice Center
Boston, MA

Investigators:
Issa J. Dahabreh, M.D., M.S.
Mei Chung, Ph.D., M.P.H.
Georgios D. Kitsios, M.D., Ph.D., M.S.
Teruhiko Terasawa, M.D., Ph.D.
Gowri Raman, M.D., M.S.
Athina Tatsioni, M.D., Ph.D.
Annette Tobar, M.D., M.S.
Joseph Lau, M.D.
Thomas A. Trikalinos, M.D., Ph.D.
Christopher H. Schmid, Ph.D.

AHRQ Publication No. 12-EHC044-EF
March 2012

Suggested citation: Dahabreh IJ, Chung M, Kitsios GD, Terasawa T, Raman G, Tatsioni A, Tobar A, Lau J, Trikalinos TA, Schmid CH. Comprehensive Overview of Methods and Reporting of Meta-Analyses of Test Accuracy. Methods Research Report. (Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I.) AHRQ Publication No. 12-EHC044-EF. Rockville, MD: Agency for Healthcare Research and Quality; March 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new health care technologies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC Program methods guidance.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality.

We welcome comments on this Methods Research Project. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
EPC Program Director and Task Order Officer
Agency for Healthcare Research and Quality

Elisabeth U. Kato, M.D., M.R.P.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Comprehensive Overview of Methods and Reporting of Meta-Analyses of Test Accuracy

Structured Abstract

Background. Medical tests play a critical role in disease screening, diagnosis, and prediction of future outcomes. Meta-analyses of diagnostic or predictive test accuracy are increasingly performed and the relevant methods are continuously evolving.

Methods. We identified systematic reviews including quantitative synthesis (meta-analysis) of test accuracy for diagnostic or predictive medical tests through MEDLINE searches (1966 to December 2009) and perusal of reference lists of eligible articles and relevant reviews. We extracted information on topics and test types covered, methods for literature synthesis and quality assessment, availability of data, and statistical analyses performed.

Results. Our searches retrieved 1,225 potentially eligible reviews of which 760 (published from 1987 to 2009) were finally considered eligible for inclusion. Eligible reviews included a median of 18 primary studies and typically examined a single index test against a single reference standard. The number of publications increased per calendar year ($P < 0.001$). Most meta-analyses pertained to cardiovascular disease (21 percent) and oncology (25 percent); the most common test categories were imaging (44 percent) and biomarker tests (28 percent). Meta-analyses used multiple electronic databases (62 percent used at least one electronic database in addition to MEDLINE; P for trend over time < 0.001) to identify eligible studies. There was a striking increase in the proportion of systematic reviews that reported assessing verification bias (P for trend < 0.001), spectrum bias (P for trend = 0.007), blinding (P for trend < 0.001), prospective study design (P for trend < 0.001), or consecutive patient recruitment (P for trend < 0.001), over time. Improvements were associated with reporting of using quality-item checklists to guide assessment of methodological quality. In statistical analyses, sensitivity (in 77 percent), specificity (in 74 percent) and diagnostic/predictive odds ratios (in 34 percent) were the most commonly used metrics. Heterogeneity tests were used in 58 percent, and subgroup or regression analyses were used in 57 percent of meta-analyses. Random effects models were employed in 57 percent of the reviews and increasingly over time (P for trend < 0.001). Theoretically motivated methods that model sensitivity and specificity simultaneously, while accounting for between-study heterogeneity, were used in a minority of reviews (11 percent) but increasingly over time (P for trend < 0.001).

Conclusion. Meta-analyses of diagnostic or predictive tests are increasingly performed. Over time there have been substantial improvements in the literature review, quality assessment and statistical analysis methods employed. Much of the improvement in quality assessment is associated with the use of quality item checklists. Advanced statistical methods have been increasingly adopted over time but their use still remains limited.

Contents

Figures

Appendixes

Background

Diagnostic and predictive tests are an important component of medical care, and clinicians rely on test results to establish diagnosis and guide patient management.[1] Despite their central role in patient care, evaluating the effectiveness of specific tests is challenging. Tests affect clinical outcomes indirectly, through the effect of test results on physicians' diagnostic thinking and subsequent management decisions, making it difficult to ascribe patient outcomes to the use of a particular test. The many existing frameworks for assessing the value of testing propose a stepwise appraisal process, moving from analytic validity (technical test performance), to clinical validity (diagnostic and predictive accuracy), clinical utility (effect on clinical outcomes) and overall cost-effectiveness assessment.[2] Primary studies that directly address all components of the assessment framework are very uncommon. Therefore, systematic reviewers are typically faced with the task of putting together the pieces of the puzzle by synthesizing studies that address each component of the framework. While the diagnostic or predictive accuracy of a medical test does not directly inform on the clinical value of testing, it is a crucial piece of the overall puzzle and one that is essential to synthesize in systematic reviews. The level of test accuracy required for tests to have any impact on clinical outcomes depends on their role (replacement, add-on, or triage test), the setting (screening, diagnosis, prognosis/prediction) of test use, and the specific clinical context.[3,4] Meta-analysis of test accuracy can provide an estimate of average test accuracy as well as identify patient-, disease- or test-related modifiers of test performance.[5]

Meta-analyses of test accuracy present particular challenges compared to reviews of randomized trials of therapeutic interventions, not only because the studies reviewed are exclusively observational, but also because of the inherent associations among the metrics of performance. Sensitivity and specificity are likely to be correlated (between studies) because of threshold effects (i.e., because changing the diagnostic threshold affects sensitivity and specificity in opposite directions), necessitating the use of multivariate analytic methods.[6] In the presence of such correlation, univariate meta-analyses of sensitivity and specificity may produce "average" values for each metric that are incompatible and have misleading confidence intervals. Only recently have these methods penetrated into common practice and into methodological guidelines aided by their implementation in readily available software.[7-11] The large number of metrics that can be used to summarize information on test accuracy has added to the analytic complexity. In addition to sensitivity and specificity, metrics such as the odds ratio,[12] area under the receiver operating characteristic (ROC) curve,[13] and likelihood ratios have been proposed for the synthesis of studies of test accuracy.[5] Finally, clinical heterogeneity is omnipresent because the studies differ so much in their settings, patient disease spectra, and versions of the tests used. This diversity often manifests as statistical heterogeneity. Thus, meta-analyses of test accuracy need to quantify and account for the presence of heterogeneity and allow the exploration of factors that may be causing it.

Early on, it was recognized that the quality of medical test accuracy studies was often inadequate.[14,15] Many items typically considered in the appraisal of studies of therapeutic interventions, such as the use of randomization, blinding of patients to the interventions used, or allocation concealment, do not apply to studies of test accuracy. A number of studies have investigated study design and reporting items that may affect estimated test accuracy, but the evidence on which items are most important is inconclusive.[16-18] Drawing on empirical evidence and expert opinion on the quality assessment of accuracy studies, the Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool was developed and published in November

2003.[19,20] This tool has now been validated for use in systematic reviews of diagnostic tests.[21] Further, a reporting checklist for primary studies of test accuracy, the Standards for the Reporting of Diagnostic Accuracy Studies (STARD),[22,23] was also published in January 2003. Although the checklist was intended as a guide for the reporting of primary research studies on diagnostic tests, the 25 STARD items pertaining to the design, analysis, and reporting of studies are often used to develop items for quality assessment in systematic reviews. As QUADAS and STARD have now been available for some time, it is reasonable to assess their impact on quality assessment methods in meta-analyses of medical tests.

Along with the overall number of meta-analytic publications, the number of meta-analyses of test accuracy studies has skyrocketed, increasing from fewer than 10 per year in the early 1990s to almost 100 publications per year in recent years. The question therefore appears to have shifted from whether meta-analysis of medical test accuracy studies is useful,[24] to what methods are best for undertaking such analyses, in terms of study identification and selection, assessment of study quality, statistical analysis and reporting. This report is the first in a series of three on meta-analysis of test accuracy, conducted by the Tufts Evidence-based Practice Center under contract with AHRQ. For this project we performed a systematic overview of meta-analyses of medical test accuracy, to assess the current state of the literature and evaluate trends over time in the methods and reporting of such studies. Here, we aimed to produce a descriptive summary of the current state of the literature of applied meta-analyses, with a focus on methods and reporting. Subsequent reports in this series will include an empirical assessment of alternative analytic methods and the development of novel methods for the analysis of diagnostic test networks.

Methods

This project reviewed meta-analyses published over two time periods: the first covered years up to 2003 and the second covered 2004-09. The search strategies, inclusion and exclusion criteria for both periods were the same. However, for studies published during the second period additional items were extracted from eligible reviews. These differences and the rationale for them are discussed in the pertinent sections below.

Search Strategy and Eligibility Criteria

We searched the MEDLINE database (1966 through to December 2009) using a combination of key words related to test accuracy and meta-analysis. The complete search strategy is presented in Appendix A.

Papers were considered eligible when they reported the findings of systematic reviews (defined as reviews using explicit methods to identify, select and extract information from primary research studies) that used quantitative synthesis (meta-analysis) methods to obtain summary estimates of diagnostic or predictive accuracy of medical tests. Our definition of tests encompassed clinical signs and symptoms. We only included English-language reviews published in full text; retrieving the full-text and extracting information from non-English articles entails substantial effort and would be unlikely to affect our conclusions. We did not consider systematic reviews that did not use quantitative analysis methods because one of our key aims was to assess the temporal evolution and current status of meta-analytic methods for synthesizing test accuracy data. We excluded reviews reporting meta-analyses based on individual patient data because they are subject to different design, analysis and reporting considerations. We also excluded Health Technology Assessment documents, evidence reports produced by the Effective Health Care Program of the Agency for Healthcare Research and Quality (AHRQ), and Cochrane Reviews of diagnostic tests; these documents are substantially longer than the typical meta-analyses published in journals and are subject to reporting conventions determined by the respective entities.

Data Extraction

Nine reviewers extracted data from nonoverlapping sets of publications in extraction forms generated using electronic data collection forms which included abbreviated operational definitions for each item. Forms were piloted using articles extracted independently by multiple reviewers and modifications were performed based on the pilot results. The final data extraction form is presented in Appendix B.

Information Extracted From Meta-Analyses Published 1966-2003

For each paper we extracted the following items: bibliographic information (first author, journal, year of publication); number of index tests, reference standard tests and the number of studies included in quantitative analyses; medical subspecialty to which tests were pertinent (cardiovascular disease, obstetrics and gynecology, gastrointestinal disease infectious disease, oncology, nephrology/urology, rheumatology, pulmonary medicine, orthopedics, psychiatry, ear-nose-throat, neurology and pediatrics); the types of test being assessed (histology/cytology/culture-based tests, clinical examination, imaging, biomarker, clinical challenge tests [e.g., pharmacological stress tests], physiologic tests [e.g., electrocardiogram, electroencephalogram] or endoscopy); details about search strategies used; quality assessment and information extracted from each primary study considered by the meta-analysis (including whether the reviews assessed blinding, spectrum bias, and verification bias in the primary studies they reviewed); use of STAndards for the Reporting of Diagnostic accuracy studies (STARD)[23]; statistical analysis (including assessment and exploration of heterogeneity, metrics used to assess test accuracy and statistical methods used for synthesizing study findings and graphically presenting these results); and assessment of comparative evidence on alternative index tests.

Information Extracted From Meta-Analyses Published 2004–2009

All data items extracted from studies published between 1966 and 2003 were also extracted from meta-analyses published between 2003 and 2009; however, from meta-analyses published during this period we extracted additional information on blinding (specifically whether index test or reference standard assessors were blinded); use of the QUADAS (first published in November 2003)[19] checklist to guide quality assessment of the primary studies; and whether the reviews collected information on the following variables from each eligible study: spectrum bias, selection criteria, number of withdrawals, number of indeterminate test results, independence of and timing of test results compared to the reference standard, and participants' sex.

During the course of data extraction, the review team met regularly to discuss specific papers, review data items, and clarify operational definitions. The majority of investigators participating in the project attended each meeting; resolutions of specific issues were reached by consensus and were circulated to all team members in writing.

Data Cleaning and Quality Control

When all reviewers completed their extractions, we merged the individual data extraction forms to generate a combined database. We queried the database to identify missing values, invalid entries (e.g., a numerical value out of the expected range) and logical inconsistencies (e.g., when a study was recorded as not using an advanced statistical method we checked that no such method was checked in the relevant fields). For every missing value identified we required the data extractor familiar with the paper to re-extract information, when necessary with the help of a second reviewer. Additionally, for continuous variables we identified entries with values differing by more than 3 standard errors from their mean value and verified them against the source documents.

All eligible meta-analyses published up to 2003 (n = 260) were extracted in duplicate. Due to the rapid increase in the number of eligible publications in more recent years, only a sample of 83 articles (17 percent of eligible studies published between 2004 and 2009) was extracted in duplicate. In all cases, discrepancies were resolved by consensus among extractors involving additional investigators, as needed (e.g., for issues relevant to the meta-analysis methods used a statistician was involved). For items that were found to be systematically different between extractors, we held discussion to ensure proper understanding of the relevant operational definitions and then, information on problematic items from all papers of the discrepant reviewers was re-extracted.

To ensure inter-reviewer consistency in the data extractions of reviews published after 2004 we implemented additional quality control measures. After data cleaning, we assessed inter-reviewer consistency by performing statistical comparisons between reviewers for all extracted variables, using chi-squared tests (for categorical variables) or analysis of variance (ANOVA) for continuous variables. To avoid the confounding effect of temporal trends, all comparisons were stratified by year of publication. Variables that reached statistical significance were considered suggestive of the existence of systematic between-reviewer differences. Each potentially problematic variable was discussed by the review team and information was re-extracted following standardization of the pertinent operational definitions.

Journal Impact Factor and Review Citation Count

We performed exploratory analyses to assess whether journal impact factor was associated with the reporting characteristics of meta-analyses and whether reporting characteristics correlated with the number of citations accrued by systematic reviews. For each eligible review we collected the citation count (on August 12th, 2011) and the 2-year impact factor of the corresponding journal (using 2010 impact factor information from the Institute of Scientific Information, ISI, Journal Citation Reports database[a]). The ISI databases do not include all MEDLINE-indexed journals, thus the total number of publications available for citation and impact factor analyses was smaller than the total number of studies included in this review (of a total of 760 systematic reviews, 732 were included in the analyses of impact factors effects and 733 were included in the analyses of citation counts; 1 study was published in a journal that is included in the databases but has not yet been assigned an impact factor). In some analyses (see below) journals were grouped into high-impact factor general medical journals versus all others. High-impact factor general medical journals were defined as the top 5 journals by 2010 impact factor in the ISI database that belong to the "Medicine, General and Internal" category (New England Journal of Medicine [NEJM], Lancet, Journal of the American Medical Association [JAMA], Annals of Internal Medicine [AIM], and Public Library of Science – Medicine [PLoS-MED]). We used these data to identify factors predicting the number of citations received by each review and to identify whether journal impact factor was associated with review reporting characteristics.

Data Analysis

We calculated descriptive statistics such as means, medians and ranges for continuous variables and proportions for categorical variables, along with appropriate measures to indicate variability around these values (standard deviations, confidence intervals or interquartile ranges). We used histograms to visualize the distributions of variables of interest and line plots to depict trends in the reporting of variables of interest over time.

We compared key methodological and reporting aspects of meta-analyses pertaining to the five most common clinical areas (cardiovascular disease, oncology, gastrointestinal disease, infectious disease, and obstetrics and gynecology) and the five most common test categories (histological/cytological/culture-based tests, aspects of the clinical examination, imaging tests, biomarkers, and physiologic tests) in our dataset. These comparisons were performed using the Fisher exact test for categorical variables or the Kruskal-Wallis test for continuous and count variables.

To detect trends over time in literature review, quality assessment, statistical analysis, and reporting characteristics of meta-analyses, we used logistic regression with each of the items of interest as the response variable and year of publication as an explanatory variable. Change in the number of studies, index and reference standard tests considered in each review were assessed using linear regression of the natural logarithm of these variables on publication year.

---

[a]2010 is the most recent year for which impact factor information is currently available in the database.

For analyses of the count of citations we used negative binomial regression with the count of citations as the dependent variable, an offset equal to the number of years since publication of each systematic review, and different reporting characteristics as explanatory variables. For analyses of the association of journal impact factor with the reporting characteristics of systematic reviews we used different reporting characteristics as binary dependent variables, and journal impact factor as an explanatory variable. We also performed analyses comparing high impact factor general medical journals versus all others. Because all analyses of citation counts and journal impact factor were exploratory and performed across multiple variables included in our database, we only report associations that reached statistical significance at the 0.001 level.

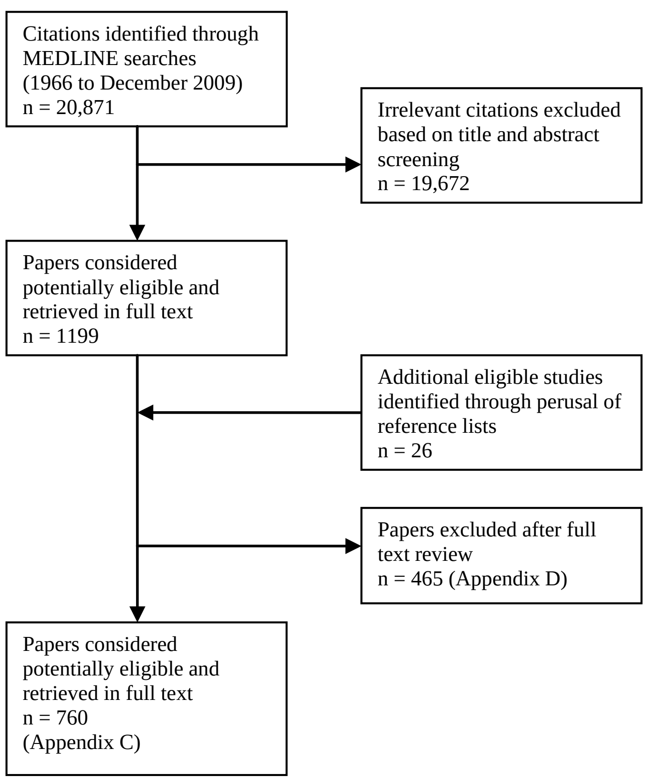
All analyses were conducted using Stata version SE/11.2 (StataCorp, College Station, TX). Statistical significance was defined as a two-sided P value < 0.05 for all comparisons (with the exception of impact factor and citation analyses).

Results

Eligible Systematic Reviews

Our searches yielded a total of 20,871 citations; after screening of titles and abstracts 1199 citations were considered potentially eligible and were retrieved in full text. An additional 26 publications were identified through perusal of reference lists of other review articles, for a total of 1225 papers reviewed in full text. Of those, 465 were excluded after full text review and 760 were considered eligible.[b] Figure 1 presents the search strategy flow. The list of included studies is presented in Appendix C; a summary of reasons for exclusion of studies reviewed in full text is presented in Appendix D. In the following sections we present a summary of the total database along with an assessment for trends over time for items of interest. Appendix E presents the regression results for trends over time for all factors assessed, both over the whole period covered (1987–2009), and for the subgroup of studies published in recent years (2005–2009); Appendix E also includes a comparison between studies published before 2003 versus those published from 2003 onwards.

Figure 1. Literature search flow



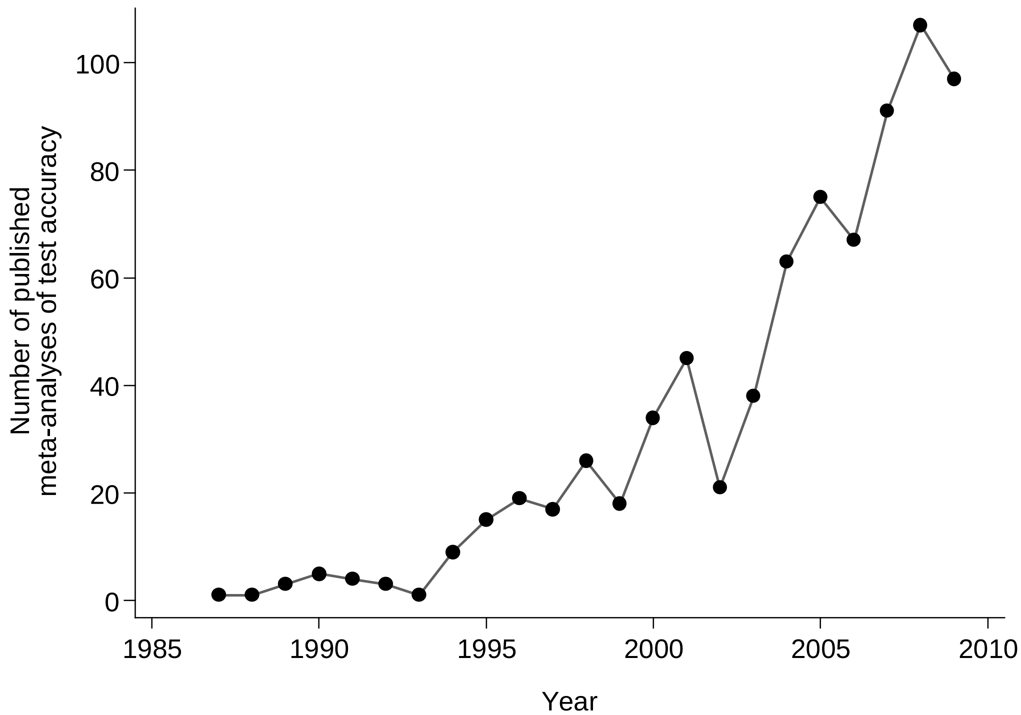
Flow of the literature search and study selection process for this review. A list of included studies is provided in Appendix C and a list of reasons for exclusion is provided in Appendix D.

---

[b]One publication[25] was a reprint of a previously published manuscript [26]; both were retained in the database.

Studies were published over more than 20 years (from 1987 to 2009), and there was a clear trend in increased number of reviews over time (P for trend < 0.001; Figure 2).

Figure 2. Number of meta-analyses of test accuracy published per year



Line plot of the number of test accuracy meta-analyses included in this overview by year of publication.

Meta-analyses had synthesized evidence from a median of 18 studies (per review), but the number of studies included varied substantially (25th–75th percentile = 11–30; minimum-maximum = 2–351). Most reviews examined imaging tests (44 percent) or molecular biomarkers (28 percent). Most tests pertained to diagnosis or prediction in oncology (25 percent), cardiology (21 percent), gastrointestinal disease (16 percent), obstetrics and gynecology (15 percent), and infectious disease (13 percent). A majority (52 percent) of meta-analyses considered a single index test and 82 percent considered a single reference standard (Table 1). Figure 3 depicts the distribution of the number of studies included in the eligible systematic reviews. Comparative analyses of two or more index tests were reported in 132 reviews (17 percent).

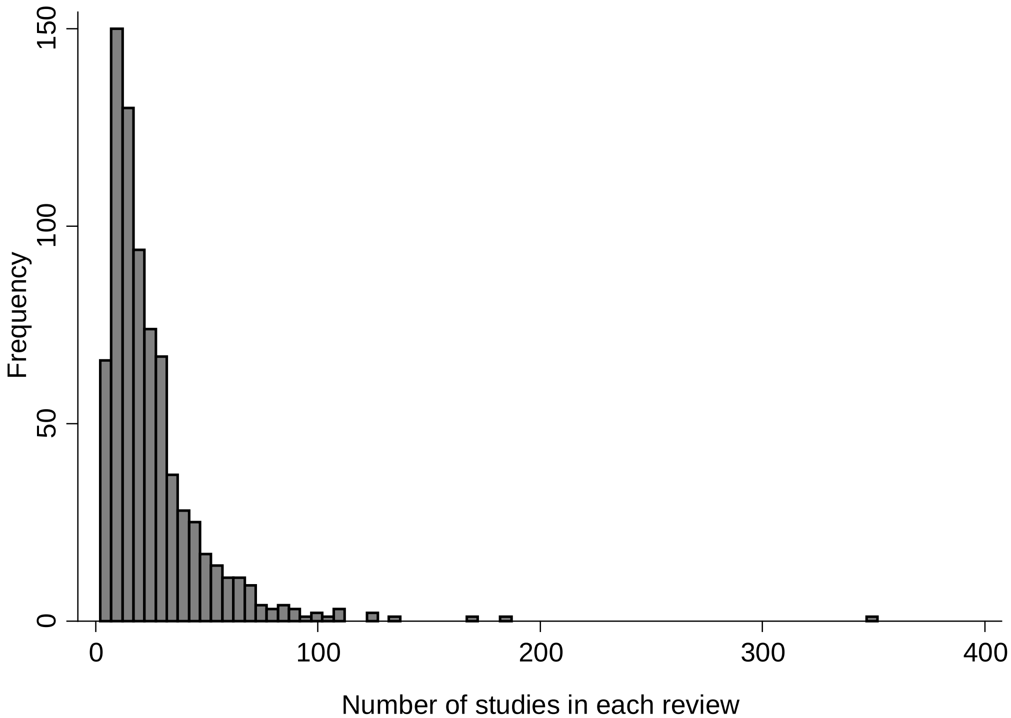
Table 1. Characteristics of eligible reviews

Characteristic	N [25th–75th percentile] (min–max)	N (%)
Topic: Oncology		188 (25)
Topic: Cardiovascular disease		160 (21)
Topic: Gastrointestinal		119 (16)
Topic: Obstetrics and gynecology		114 (15)
Topic: Infectious disease		98 (13)
Topic: Pulmonary medicine		68 (9)
Topic: Orthopedics		44 (6)
Topic: Nephrology and urology		38 (5)
Topic: Neurology		37 (5)
Topic: Pediatrics		29 (4)
Topic: Psychiatry		22 (3)
Topic: Ear-nose-throat		19 (3)
Topic: Rheumatology		8 (1)
Test type: Imaging		336 (44)
Test type: Biomarker		211 (28)
Test type: Clinical exam		112 (15)
Test type: Histology, cytology, or culture		103 (14)
Test type: Physiologic test		40 (5)
Test type: Challenge/stress test		31 (4)
Test type: Endoscopic examinations		21 (3)
Index test: Per review, median	1 [1–3] (1–56)	
Reviews with a single index test		396 (52)
Reviews with 2 index tests		157 (21)
Reviews with 3 index tests		68 (9)
Reviews with 4 index tests		43 (6)
Reviews with ≥5 index tests		96 (13)
Reference standard tests per review, median	1 [1–1] (1–7)	
Reviews with a single reference standard test		625 (82)
Reviews with 2 reference standard tests		74 (10)
Reviews with 3 reference standard tests		38 (5)
Reviews with 4 reference standard tests		14 (2)
Reviews with ≥5 reference standard tests		9 (1)
Included studies: Per review, median	18 [11–30] (2–351)	
Reviews with 2–10 studies		180 (24)
Reviews with 11–20 studies		246 (32)
Reviews with 21–30 studies		151 (20)
Reviews with 31–40 studies		63 (8)
Reviews with 41–50 studies		45 (6)
Reviews with ≥51 studies		75 (10)
Publications 1985–1989		5 (<1)
Publications 1990–1994		22 (3)
Publications 1995–1999		95 (13)
Publications 2000–2004		201 (26)
Publications 2005–2009		437 (58)

max = maximum; min = minimum.

“Clinical exam” denotes the assessment of aspects of the clinical examination as diagnostic tests; “challenge/stress test” denotes tests such as the glucose challenge test for diabetes, or stress tests (pharmacological or activity-based). The percentages of medical topics and test types do not sum up to 100% because many test uses could be classified under multiple topics and some reviews assessed more than one test types.

Figure 3. Number of primary studies included in each meta-analysis of test accuracy

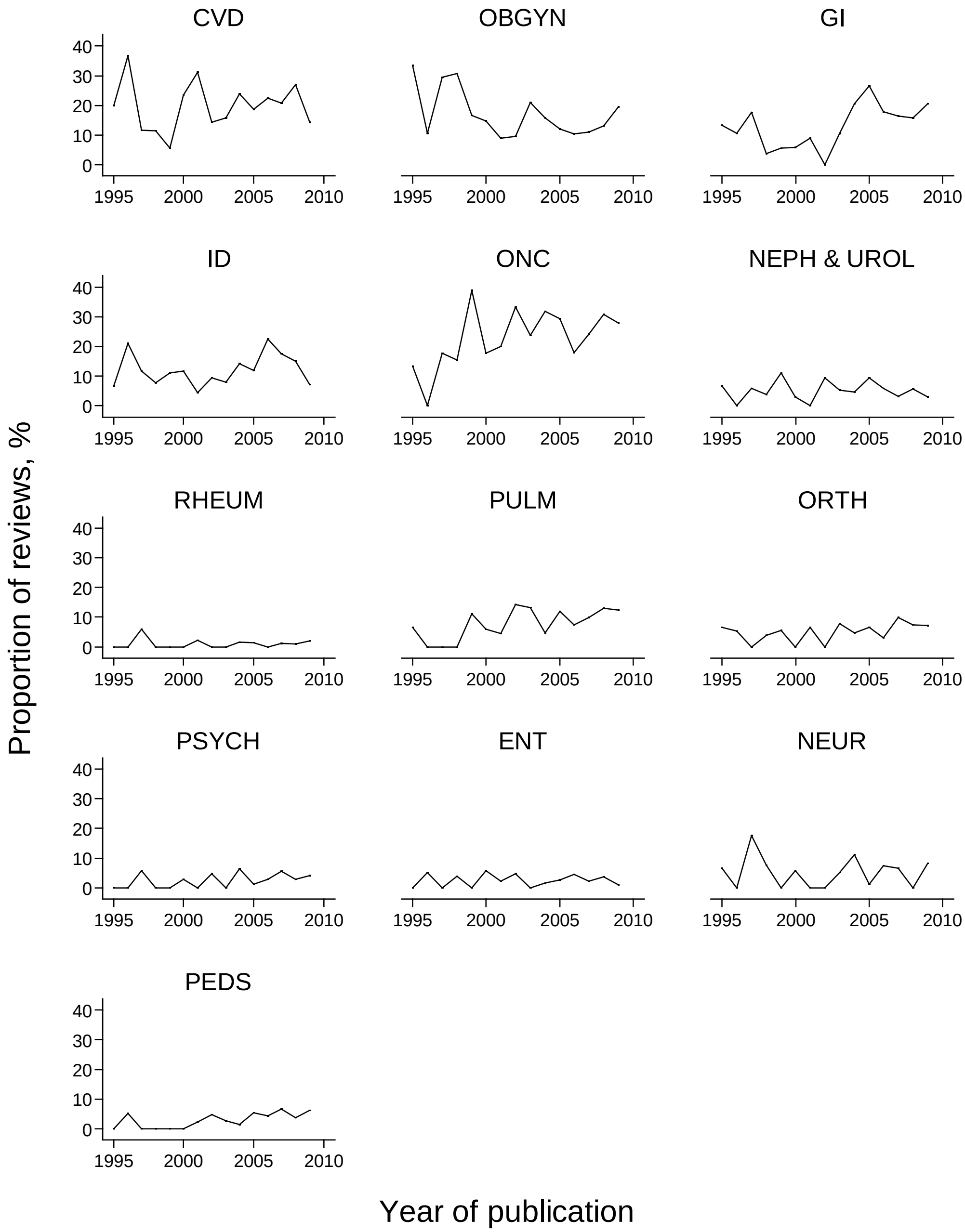


Histogram of the number of primary studies included in each meta-analysis of test accuracy included in this overview. Bin width was set at 5 studies.

Over time, there was no statistically significant change in the number of studies included in each review (-1 percent per year; 95 percent CI -2, 0.3; P = 0.147), or the number of reference standards considered (0 percent per year; 95 percent CI -0.6, 0.7; P = 0.902). However, there has been a small reduction in the number of index tests assessed (-1.2 percent per year; 95 percent CI -2.4, 0; P = 0.040). The proportion of reviews reporting comparative analyses of at least two index tests has not changed over time (per year odds ratio, OR = 0.98; 95 percent CI 0.94, 1.02; P = 0.292).

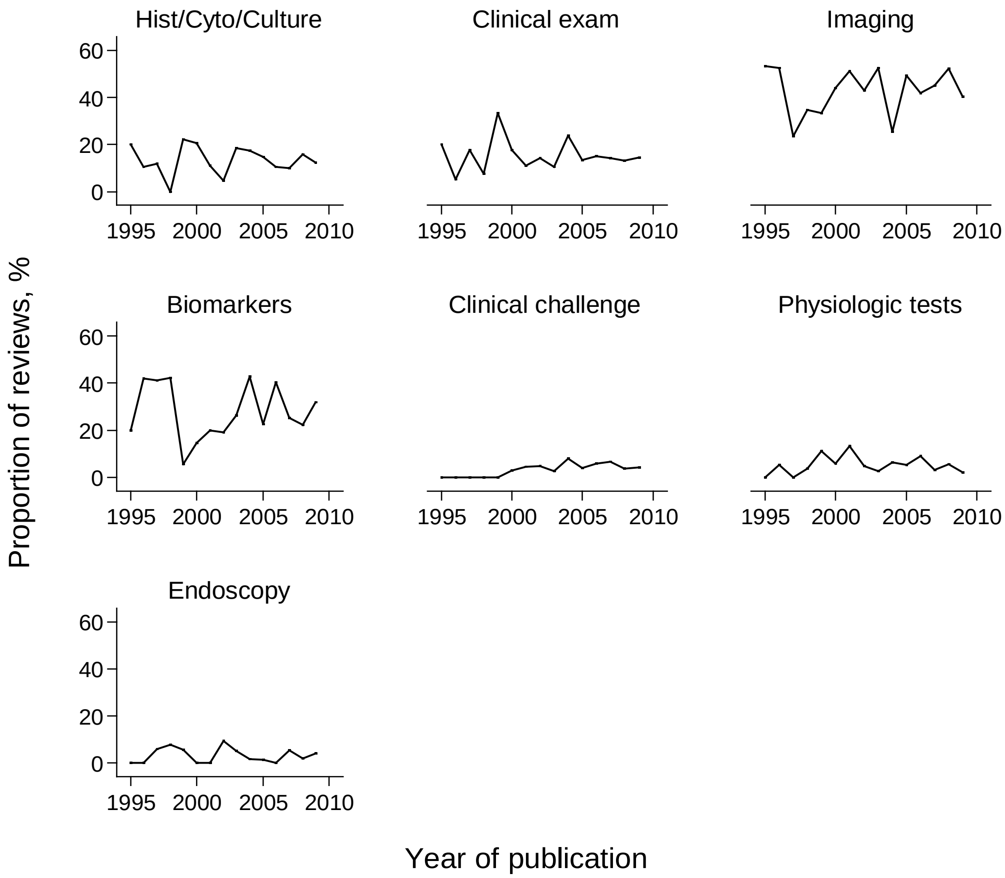
Over time there was an increase in meta-analyses pertaining to gastrointestinal disease (per year OR = 1.07; 95 percent CI 1.02, 1.13; P = 0.008), oncology (per year OR = 1.05; 95 percent CI 1.01, 1.09; P = 0.020), orthopedics (per year OR = 1.09; 95 percent CI 1.00, 1.19; P = 0.041) and pulmonary medicine (per year OR = 1.10; 95 percent CI 1.03, 1.18; P = 0.005). Changes over time were non-significant for other clinical topics (Figure 4). There was also a borderline increase in the proportion of meta-analyses assessing clinical challenge tests (per year OR = 1.11; 95 percent CI 1.00, 1.23; P = 0.046) (Figure 5).

Figure 4. Trends over time in the proportion of meta-analyses, by clinical field



Line plot of the annual proportion of meta-analyses of test accuracy in various clinical fields.
CVD = cardiovascular disease; OBGYN = obstetrics and gynecology; GI = gastrointestinal disease; ID = infectious disease; ONC = oncology; NEPH & UROL = nephrology and urology; RHEUM = rheumatology; PULM = pulmonary medicine; ORTH = orthopedics; PSYCH = psychiatry; ENT = ear-nose-throat; NEUR = neurology; PEDS = pediatrics. Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Figure 5. Trends over time in the proportion of meta-analyses, by test category



Line plot of the annual proportion of meta-analyses of test accuracy in various test categories.
Hist/Cyto/Culture = histological, cytological, or culture-based tests; clinical challenge = clinical challenge tests
Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Literature Review Methods in Reviews of Test Accuracy

Most reviews reported searching multiple electronic databases. MEDLINE searches were nearly universal (96 percent of all reviews) and 62 percent of the reviews reported searching at least one electronic databases in addition to MEDLINE. Searches of Embase (47 percent) and the Cochrane Library (30 percent) were also common. References lists of eligible studies and relevant review articles were also considered in a large proportion of the reviews (76 percent and 28 percent, respectively). On the contrary, contacting experts in the field (17 percent) and obtaining unpublished information (12 percent) were less common.

To guide the selection of eligible studies, 19 percent of eligible reviews reported using quality criteria and 21 percent reported using a minimum cut-off sample size. These cut-offs were generally low (median = 10 participants), but some reviews excluded studies of even moderate sample size (25^{th}–75^{th} percentile = 10–20; 99^{th} percentile = 100). Reviews often considered only studies published in English (36 percent); however, 31 percent of reviews explicitly reported not imposing any language restrictions.

Table 2 summarizes information on the databases searched, the reporting of search strategies in the eligible reviews, and the study selection characteristics of eligible reviews.

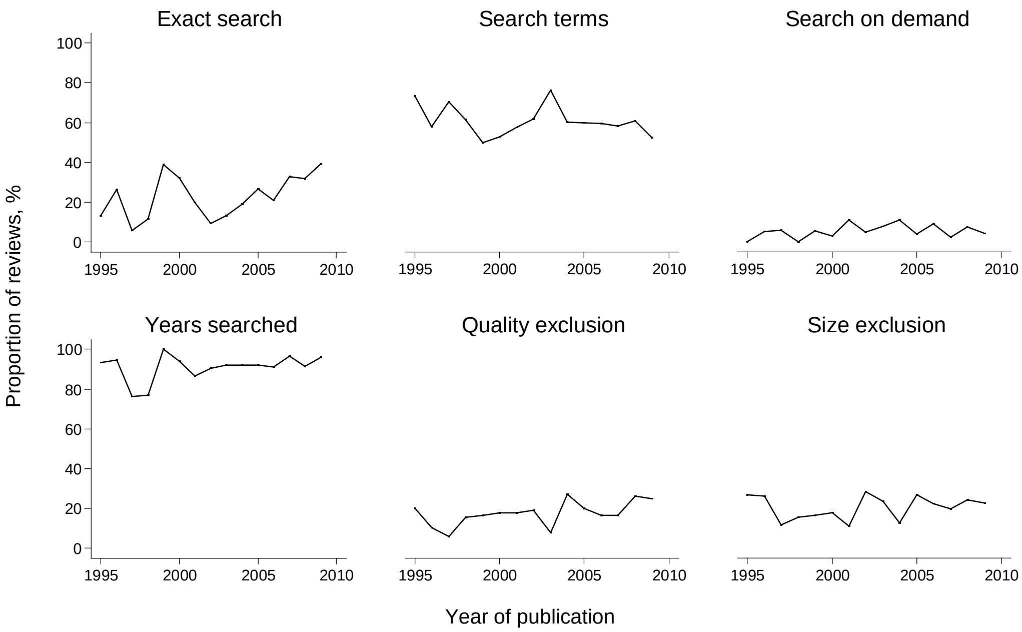
Table 2. Literature search and study selection methods employed in reviews of test accuracy

Characteristic	Number [25th–75th percentile] (min–max)	Number (%)
Availability of the search strategy: Reporting of the exact search string		195 (26)
Availability of the search strategy: Reporting of search terms		445 (59)
Availability of the search strategy: Search strategy available upon request		43 (6)
Availability of the search strategy: No information reported		77 (10)
Reporting of the years searched		697 (92)
Exclusion of studies based on quality		146 (19)
Exclusion of studies based on sample size		159 (21)
Median sample size cut-off	10 [10–20] (4–1000)	
English-only (or other single language)		271 (36)
English and other specific languages		90 (12)
No language restrictions		235 (31)
No information on selection based on language		164 (22)
Database: MEDLINE		729 (96)
Database: Embase		358 (47)
Database: Cochrane library		228 (30)
Database: CINAHL		82 (11)
Database: ISI WOK/ SCI		74 (10)
Database: Current Contents		35 (5)
Database: Other specific database		202 (27)
Source of eligible studies: Conference proceedings		88 (12)
Source of eligible studies: Bibliographies of eligible articles		576 (76)
Source of eligible studies: Bibliographies of relevant review articles		213 (28)
Source of eligible studies: Experts in relevant fields		132 (17)
Source of eligible studies: Manual/electronic searches of specific journals		105 (14)
Source of eligible studies: Source of eligible studies: Manufacturers of tests/assays		33 (4)
Source of eligible studies: Unpublished information		93 (12)

CINAHL = Cumulative Index to Nursing and Allied Health Literature; ISI WOK/ SCI = Institute of Scientific Information Web of Knowledge/Science Citation Index; max = maximum; min = minimum
Unless stated otherwise, the total sample size for analyses in this table is 744. A single review considering only Korean-language articles is included here along with the “English-only” reviews, to indicate the consideration of articles in a single language only. Percentages may not sum to 100 percent due to rounding.

Figure 6 presents trends over time in the reporting of search strategies and methods for study selection for inclusion in meta-analysis. Over time there has been a substantial increase in the number of reviews reporting the exact search strategy used (per year OR = 1.09; 95 percent CI 1.05, 1.14; $P < 0.001$) and the years searched (per year OR = 1.08; 95 percent CI 1.02, 1.13; $P = 0.004$). However, there has also been an increase in the number of reviews using quality criteria to select studies for inclusion (per year OR = 1.06; 95 percent CI 1.02, 1.11; $P = 0.007$).

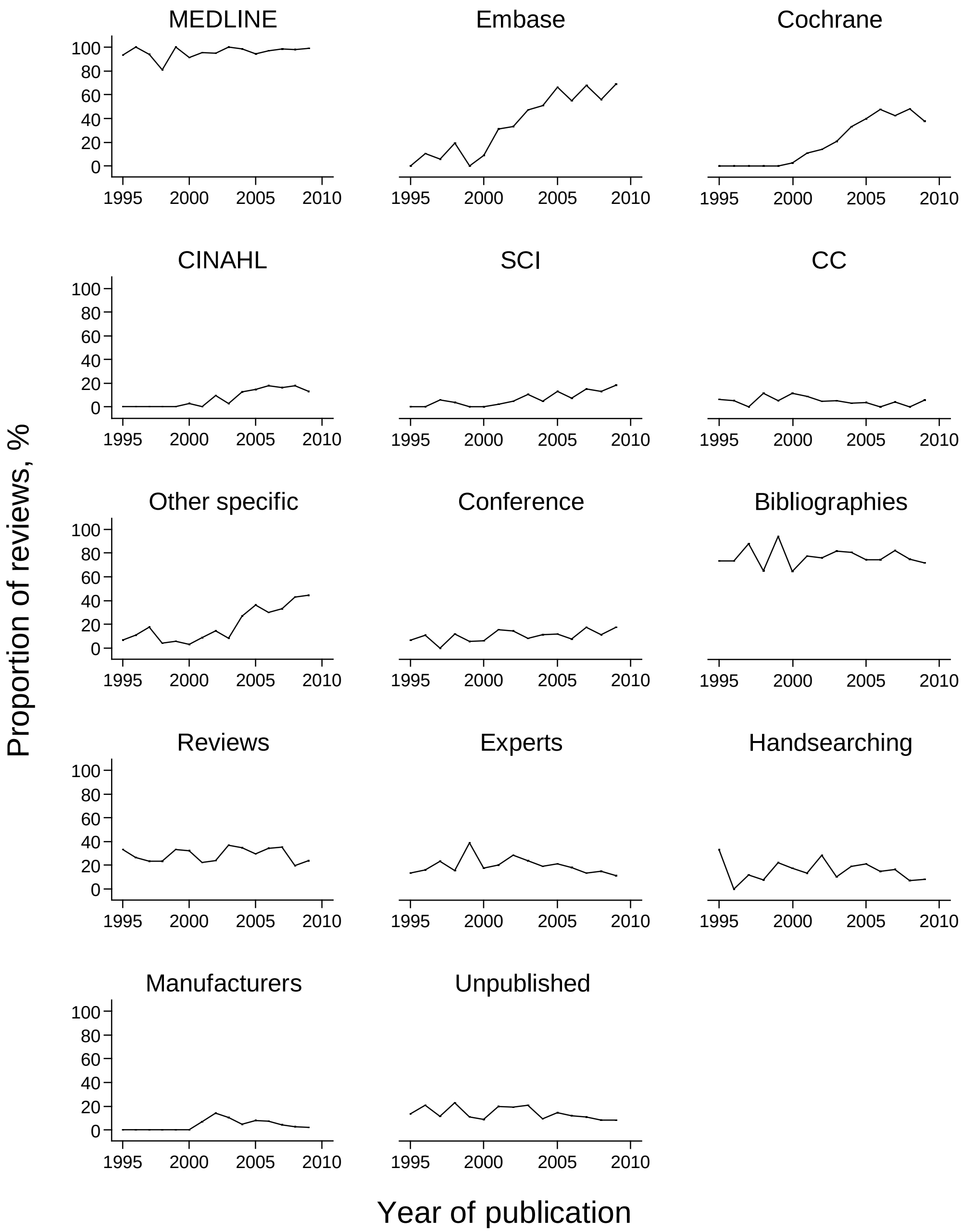
Figure 6. Trends over time in the reporting of search strategies and study selection criteria



Line plots of the annual proportion of meta-analyses of test accuracy using specific reporting practices for their search strategies or study selection criteria.
Exact search = reporting of the search strategy in a way that can be directly replicated; Search terms = reporting of search terms without the full search strategy or Boolean operands; Search on demand = the full search strategy is available from the authors or in a website (other than that of the journal publishing the meta-analysis); years searched = reporting of the years covered by the searches; quality exclusion = exclusion of studies from the meta-analysis based on quality criteria; size exclusion = exclusion of studies from the meta-analysis based on sample size criteria.
Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Figure 7 presents trends over time in the use of electronic databases and other literature sources by systematic reviews of test accuracy. Overall, there was an increasing trend in the use of electronic databases other than MEDLINE (per year OR = 1.30; 95 percent CI 1.24, 1.35; $P < 0.001$); this appeared to be due to the increasing use of Embase (per year OR = 1.28; 95 percent CI 1.23, 1.35; $P < 0.001$); the Cochrane libraries (per year OR = 1.30; 95 percent CI 1.23, 1.38; $P < 0.001$); the Science Citation Index (or other ISI databases, per year OR = 1.18; 95 percent CI 1.09, 1.28; $P < 0.001$); CINAHL (per year OR = 1.25; 95 percent CI 1.15, 1.36; $P < 0.001$); or other specific electronic databases (per year OR = 1.25; 95 percent CI 1.18, 1.32; $P < 0.001$).

Figure 7. Trends over time in the proportion of meta-analyses using specific databases or other sources to identify eligible studies

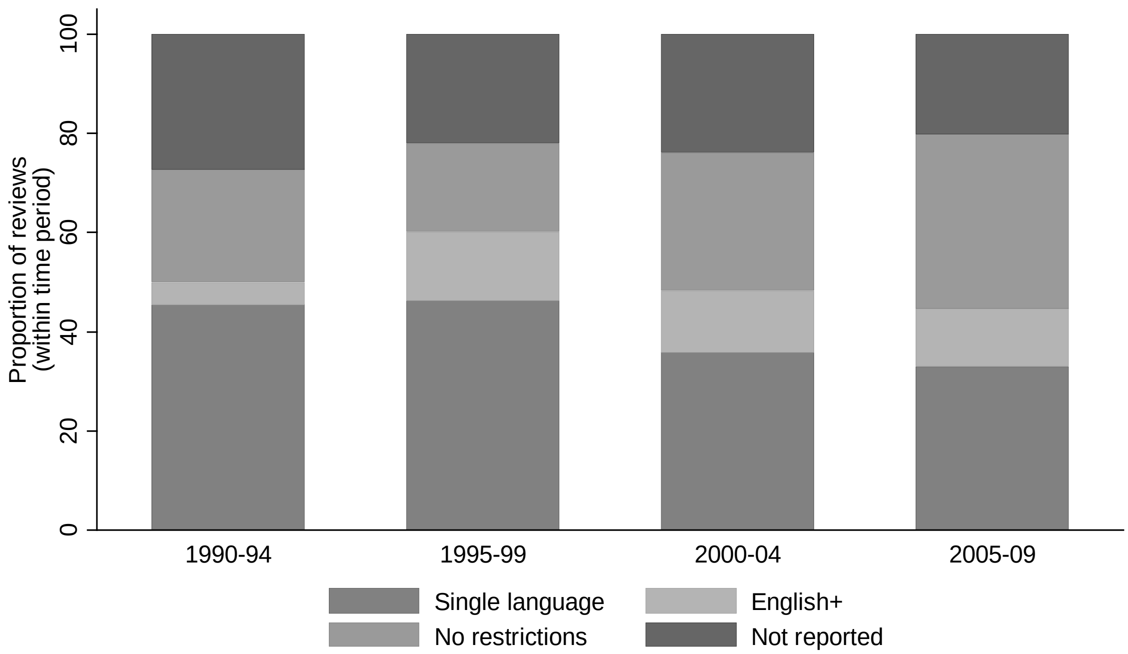


Line plots of the annual proportion of meta-analyses of test accuracy using specific databases or other sources to identify eligible studies. Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Bibliographies = perusal of reference lists of included studies; CC = current contents; CINAHL = Cumulative Index to Nursing and Allied Health Literature; Cochrane = searching of databases maintained by the Cochrane Collaboration; Experts in the field consulted to provide additional studies; Handsearching = searching manually (or electronically) the contents of selected journals; Other specific = searching of other specific electronic databases; Reviews = perusal of the reference lists of relevant review articles; SCI = Science Citation Index or other Institute of Scientific Information databases; Unpublished = search for studies not published in the peer-reviewed literature.

Figure 8 presents trends over time in the handling of languages other than English in reviews of test accuracy. There has been an increase in the proportion of studies that explicitly reported considering non-English language articles (considering studies published in at least one language other than English, per year OR = 1.05; 95 percent CI 1.02, 1.09; P = 0.003). This increase is mostly due to an increasing number of reviews not imposing any language restrictions. As expected, there has been a concomitant decrease in reviews considering English language studies only (per year OR = 0.96; 95 percent CI 0.93, 0.99; P = 0.028).

Figure 8. Trends over time in the proportion of studies searching for or considering for inclusion studies based on language of publication criteria



Stacked bar graph of the annual proportion of meta-analyses of test accuracy using different language criteria to identify eligible studies.

Single language = reviews only considering studies published in a single language (in all but one case the language used was English, a single review considered only studies published in Korean); English+ = reviews considering studies published in English and at least one more specific Language; No restrictions = reviews not using language restrictions to select eligible studies; Not reported = studies not providing information on the languages considered.

Results are shown only after 1990 because the number of meta-analyses in previous years was too small and proportions were unstable.

Factors Affecting the Number of Studies Included in Each Review

The number of studies included in each meta-analysis did not appear to be affected by the search of electronic databases in addition to MEDLINE (0.2 percent difference between reviews that included at least one additional electronic database; 95 percent CI -12.7, 12.1 percent; P = 0.967) or the inclusion of languages other than English (8.3 percent more studies in reviews including non-English language studies; 95 percent CI -2.7, 21 percent; P = 0.139), after adjusting for publication year, clinical topic, and test category.

Quality Assessment and Use of Checklists

The majority of reviews performed some quality assessment of the studies they included. This assessment was based on the QUADAS checklist in 20 percent (27 percent after 2004) of reviews. Nine percent (12 percent after 2003) of reviews reported using the STARD guideline to develop items for quality assessment of the primary studies they included. Commonly assessed items included blinding (65 percent), prospective recruitment of patients (59 percent), verification bias (48 percent), and the description of the reference standard used (88 percent). Blinding of test assessors was examined in more detail for articles published since 2004: blinding of the index test assessor to the reference standard results was reported in 53 percent, and blinding of the reference standard assessor to the index test results in 50 percent of the 500 studies published between 2004 and 2009 (Table 3).

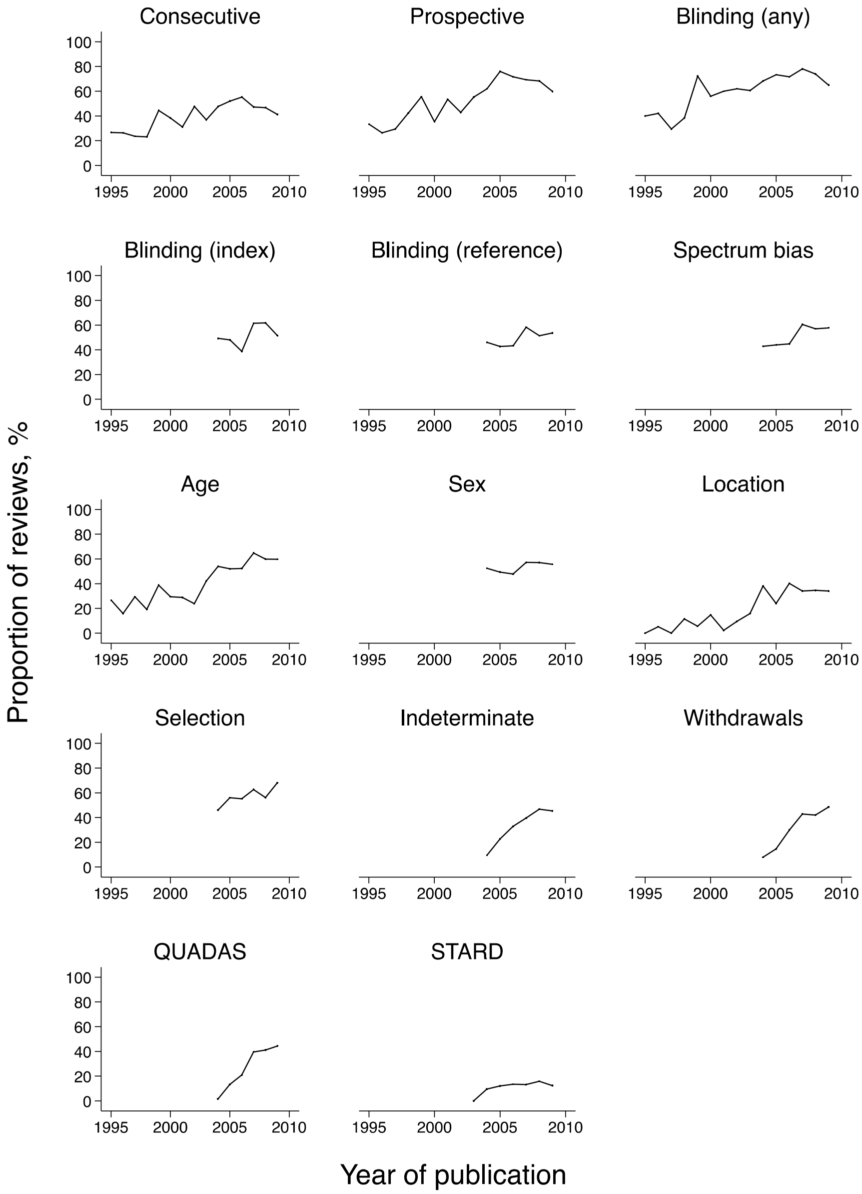
Table 3. Quality assessment in systematic reviews of medical test accuracy studies

Characteristic	Studies, n (%)
Settings of test use	295 (39)
Recruitment of consecutive patients	322 (42)
Prospective/retrospective design	446 (59)
Exact description of the reference standard	667 (88)
Expertise of the test readers	116 (15)
Any blinding (all eligible studies)	496 (65)
Blinding assessors of the index test to the reference standard (500 studies published since 2004)	265 (53)
Blinding assessors of the reference standard test to the index test (500 studies published since 2004)	250 (50)
Blinding unspecified (500 studies published since 2004)	74 (15)
No blinding (500 studies published since 2004)	141 (28)
Demographic characteristics of participants: Age	361 (48)
Demographic characteristics of participants: Sex (500 studies published since 2004)	269 (54)
Location of studies	190 (25)
Spectrum bias (500 studies published since 2004)	262 (52)
Selection bias (500 studies published since 2004)	291 (58)
Time between the performance of the index and reference standard tests (500 studies published since 2004)	201 (40)
Test independence (500 studies published since 2004)	190 (38)
Indeterminate results (500 studies published since 2004)	175 (35)
Withdrawals (500 studies published since 2004)	167 (33)
Verification bias	367 (48)
Quality assessment based on the QUADAS tool	148 (19)
Quality assessment based on use of items from the STARD reporting guideline	65 (9)

QUADAS = Quality Assessment of Diagnostic Accuracy Studies; STARD = Standards for Reporting of Diagnostic Accuracy.

Over time, assessment of specific quality items has generally increased. (Figure 9). For example, there has been an increase in the number of reviews appraising verification bias (per year OR = 1.17; 95 percent CI 1.13, 1.22; $P < 0.001$); spectrum bias (per year OR = 1.16; 95 percent CI 1.04, 1.29; $P = 0.007$); test assessor blinding (per year OR = 1.09; 95 percent CI 1.05, 1.13; $P < 0.001$); whether study design was prospective (per year OR = 1.12; 95 percent CI 1.09, 1.16; $P < 0.001$); and whether patients were recruited consecutively (per year OR = 1.07; 95 percent CI 1.04, 1.11; $P < 0.001$).

Figure 9. Trends over time in the proportion of meta-analyses of test accuracy appraising specific quality items or using quality assessment checklists



Line plot of the annual proportion of meta-analyses of test accuracy appraising selected quality items among the primary studies they included. Information for the following variables was only collected for studies published 2004 onwards: Blinding (index), Blinding (reference), Spectrum bias, Sex, Selection bias, Indeterminate test results, and Withdrawals. The QUADAS tool and the STARD checklist were published in November 2003 and January 2003, respectively. Results for all variables are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Consecutive = reviews assessing whether the primary studies enrolled patients consecutively; Prospective = reviews assessing whether the primary studies had a prospective design; Blinding = any assessment of blinding in the primary studies; Blinding (index) = assessment of whether the index test assessor was blinded to the reference standard results; Blinding (reference) = assessment of whether the reference standard assessor was blinded to the index test results; Age = reviews that extracted information on participant age from the primary studies; Sex = reviews that extracted information on participant sex from the primary studies; Location = reviews that extracted information on primary study locations; Selection = reviews that extracted information on the selection criteria of the primary studies; Indeterminate = reviews that examined the handling of indeterminate test results in the primary studies; Withdrawals = reviews that examined the handling of withdrawals in the primary studies; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; STARD = Standards for Reporting of Diagnostic Accuracy.

These improvements in recent years in the quality assessment of primary studies have been associated with the increasing use of checklists of items relevant to study quality (Table 4): since 2004, the QUADAS tool has increasingly been used to guide quality assessment (per year OR = 1.62; 95 percent CI 1.41, 1.86; P < 0.001). Similarly, since 2003, the STARD reporting guideline has also been increasingly (but not quite statistically significantly) used to develop items for the quality assessment of primary studies of test accuracy (per year OR = 1.15; 95 percent CI 0.99, 1.33; P = 0.059).

Table 4. Use of QUADAS or STARD to guide quality assessment in reviews of medical tests (2004–2009)

Year	% of reviews using QUADAS	% of reviews using STARD
2004	1.6	9.5
2005	13.3	12.0
2006	20.9	13.4
2007	39.6	13.2
2008	41.1	15.9
2009	44.3	12.4

QUADAS = Quality Assessment of Diagnostic Accuracy Studies; STARD = Standards for Reporting of Diagnostic Accuracy

Statistical Analyses and Presentation of Results in Reviews of Test Accuracy

The most popular test accuracy metrics used in meta-analysis were sensitivity (77 percent) and specificity (74 percent). Diagnostic odds ratios (34 percent) and likelihood ratios (31 percent) were also commonly used (Table 5). Quantitative results were often presented in forest plots (39 percent) or ROC curves (53 percent); other graphical displays were uncommon. Heterogeneity tests were performed in 58 percent of the available studies; potential causes of underlying heterogeneity were explored in 57 percent of analyses (33 percent using exclusively subgroup analyses exclusively and 24 percent using meta-regression with or without subgroup analyses). The most commonly used heterogeneity metrics were Cochran's Q statistic, Fisher's exact test, and the I^2 index. Random effects models were used in the majority (57 percent) of the studies.

Table 5. Statistical analyses and presentation of results in reviews of test accuracy

Characteristics	Studies, n (%)
Use of random effects models	436 (57)
Metrics used in quantitative analyses: Sensitivity	582 (77)
Metrics used in quantitative analyses: Specificity	560 (74)
Metrics used in quantitative analyses: OR	257 (34)
Metrics used in quantitative analyses: Likelihood ratios	236 (31)
Metrics used in quantitative analyses: Predictive values	97 (13)
Metrics used in quantitative analyses: Accuracy	42 (6)
Metrics used in quantitative analyses: AUC	40 (5)
Metrics used in quantitative analyses: Q*	26 (3)
Display of synthesis results: Forest plots	300 (39)
Display of synthesis results: ROC curves	403 (53)
Heterogeneity testing	439 (58)
Exploration of heterogeneity: Meta-regression analyses (+/- subgroup analyses)	180 (24)
Exploration of heterogeneity: Subgroup analyses only	247 (33)
No exploration of heterogeneity	333 (44)
Statistical analyses: Univariate meta-analysis	660 (87)

Table 5. Statistical analyses and presentation of results in reviews of test accuracy (continued)

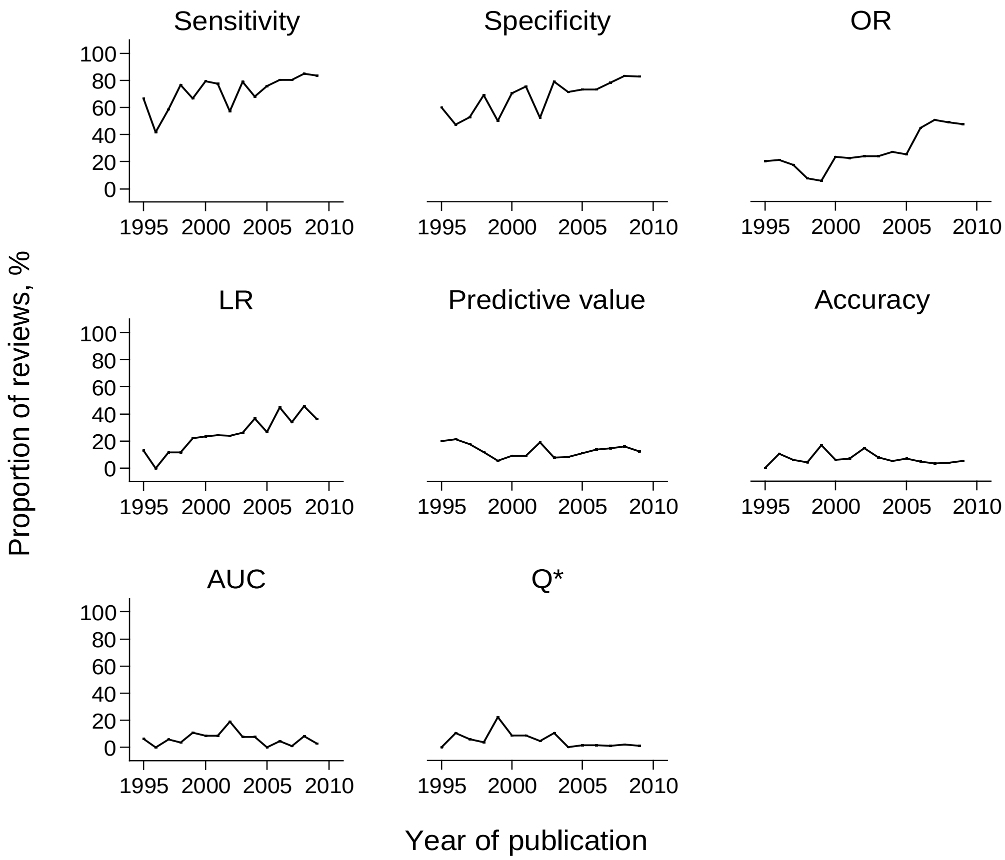
Characteristics	Studies, n (%)
Statistical analyses: ROC-based methods (including sROC and hsROC)	379 (50)
Statistical analyses: Advanced statistical methods – BREM/hsROC	70 (9)
Statistical analyses: Other advanced models	11 (1)
sROC model (only among studies using ROC analyses): Moses-Littenberg	326 (86)
sROC model (only among studies using ROC analyses): Rutter-Gatsonis	24 (6)
sROC model (only among studies using ROC analyses): Other	29 (7)
Comparative analyses of index tests	132 (17)
Direct comparative analyses of index tests (only among the 131 studies reporting on test comparisons)	33 (25)
Indirect comparative analyses of index tests (only among the 131 studies reporting on test comparisons)	98 (75)
Bayesian statistical analyses	17 (2)
Reporting of data to replicate analyses: Counts reported or can be calculated	448 (59)
Reporting of data to replicate analyses: Data not available	312 (41)

AUC = area under the curve; BREM = bivariate random effects meta-analysis; OR = diagnostic/predictive odds ratio; HsROC = hierarchical summary receiver operating characteristic; ROC = receiver operating characteristic; sROC = summary receiver operating characteristic

Statistical analyses most often used univariate (one outcome at a time) meta-analyses (87 percent) and the fixed effects summary receiver operating curve characteristics method as described by Moses and Littenberg[13,27] (86 percent of the studies performing ROC analyses). More theoretically motivated methods, such as bivariate random effects[6,28] or hierarchical summary ROC curve models[9,10], were rarely used (11 percent), although this is changing (see below for time trend).

Figure 10 presents trends over time in the proportion of studies using specific metrics for meta-analysis of test accuracy information. Over time there has been increasing use of the diagnostic OR (per year OR = 1.17; 95 percent CI 1.12, 1.22; $P < 0.001$), sensitivity (per year OR = 1.07; 95 percent CI 1.03, 1.11; $P < 0.001$), specificity (per year OR = 1.08; 95 percent CI 1.04, 1.12; $P < 0.001$), and likelihood ratios (per year OR = 1.13; 95 percent CI 1.08, 1.18; $P < 0.001$), as metrics for meta-analyses of test accuracy.

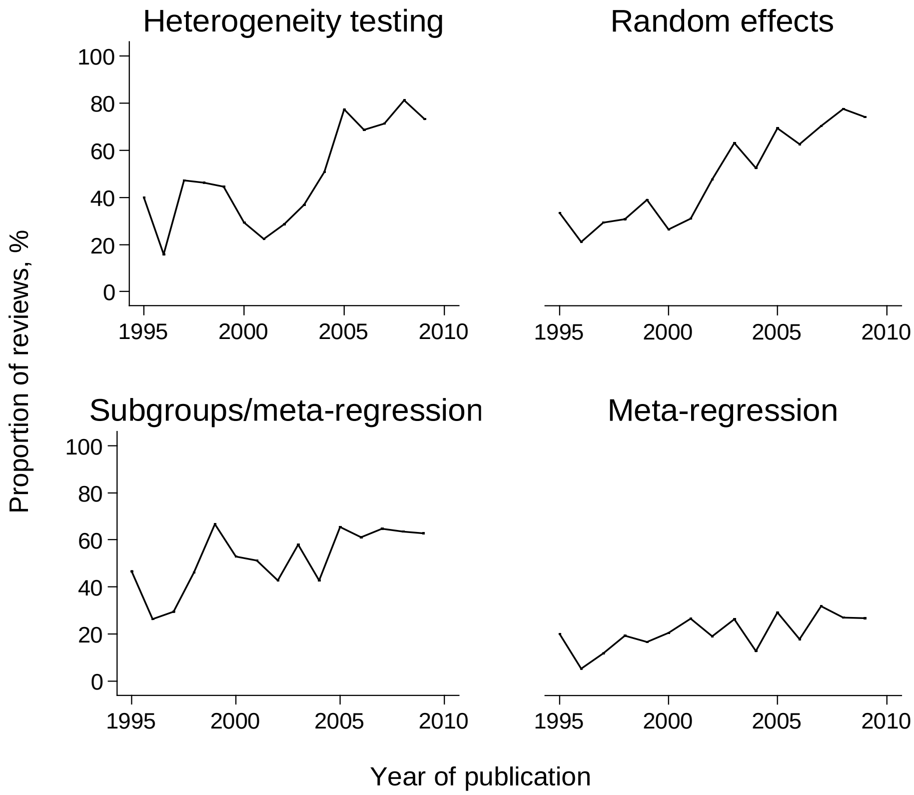
Figure 10. Trends over time in the proportion of meta-analyses using each metric of test accuracy



Line plot of the annual proportion of meta-analyses of test accuracy using each metric for quantitative evidence synthesis. OR = odds ratio; LR = likelihood ratio; AUC = area under the receiver operating characteristic curve; Q* = point where sensitivity equals specificity on SROC curve). Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Figure 11 presents trends over time in the proportion of reviews assessing, accounting for, and exploring heterogeneity. There has been a clear increase in the number of studies assessing heterogeneity using statistical tests (per year OR = 1.21; 95 percent CI 1.17, 1.26; P < 0.001), and exploring the underlying reasons leading to heterogeneity using subgroup or meta-regression methods (per year OR = 1.08; 95 percent CI 1.04, 1.12; P < 0.001). This increase has been mostly due to the use of subgroup analyses, as the proportion of reviews performing meta-regression analyses has not changed significantly over time; per year OR = 1.03 (95 percent CI 0.95, 1.11; P = 0.484). Use of random effects models also increased over time (per year OR = 1.21; 95 percent CI 1.16, 1.26; P < 0.001).

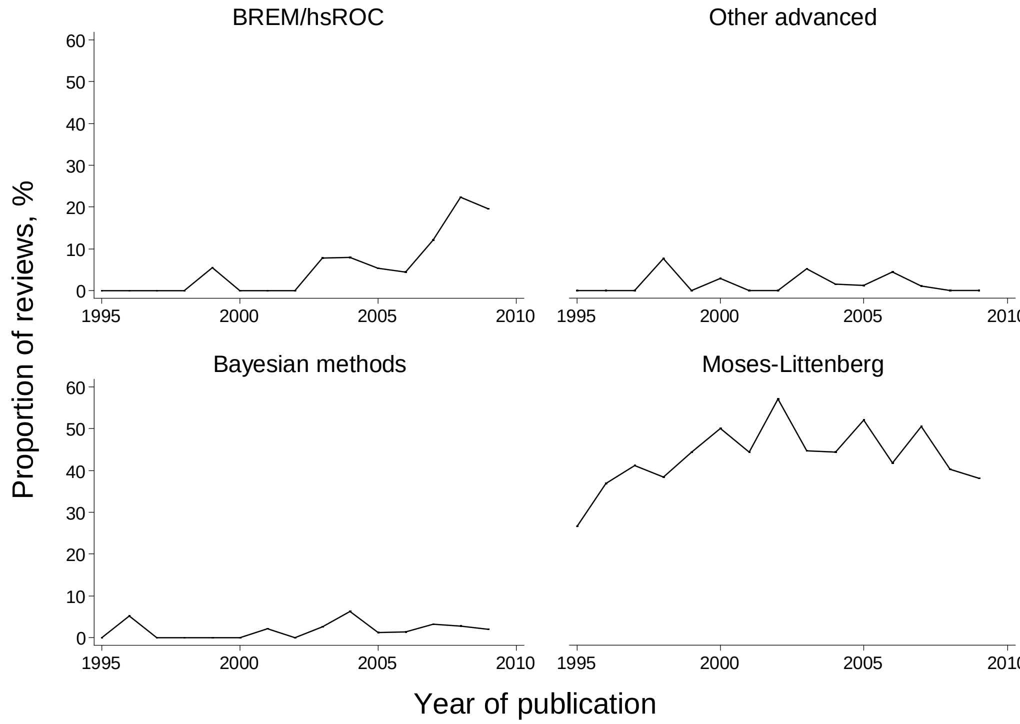
Figure 11. Trends over time in the proportion of meta-analyses assessing, accounting for in the analyses and exploring heterogeneity



Line plot of the annual proportion of meta-analyses of test accuracy using tests for heterogeneity, employing random effects meta-analysis models, exploring heterogeneity (using subgroup or regression analyses), and using meta-regression analyses. Results are shown only after 1995 because the number of meta-analyses in previous years was too small and proportions were unstable.

Figure 12 presents trends in the proportion of studies using advanced statistical methods (for comparison a trend line for studies using the Moses-Littenberg method is also presented). Overall, the proportion of studies using advanced meta-analysis methods has increased over time (per year OR = 1.27; 95 percent CI 1.16, 1.39; P < 0.001). Specifically, the bivariate random effects meta-analysis model and the hierarchical sROC model are increasingly used: per year OR = 1.42; 95 percent CI 1.26, 1.60; P < 0.001) (Table 6). These methods were used about 20 percent of the time in 2008 and 2009 following publication of several articles that recommended their use.[6,28-30] Bayesian statistical methods have been rarely utilized.

Figure 12. Trends over time in the proportion of meta-analyses using advanced statistical methods for quantitative evidence synthesis



Line plot of the annual proportion of meta-analyses of test accuracy using advanced meta-analysis methods, such as the bivariate random effects meta-analysis or the hierarchical summary receiver operating characteristic curve models; other advanced methods (in most cases, random effects variants of the Moses-Littenberg sROC method); or Bayesian analysis methods. Results are shown only for years after 1995 because the statistical methods of interest were practically not used at all during earlier years and the number of meta-analyses was relatively small leading to instability of the estimated annual proportions. The plot for meta-analyses using the Moses-Littenberg method is presented for comparison, BREM = bivariate random effects meta-analysis; hsROC = hierarchical receiver operating characteristic curve; sROC = summary receiver operating characteristic method.

Table 6. Use of advanced statistical methods in reviews of medical tests (2000-09)

Year	% of Reviews Using BREM/hsROC
2000	0.0
2001	0.0
2002	0.0
2003	7.9
2004	7.9
2005	5.3
2006	4.5
2007	12.1
2008	22.4
2009	19.6

BREM = bivariate random effects meta-analysis; hsROC = hierarchical receiver operating characteristic curve

Comparison of Reviews in the Five Most Commonly Assessed Medical Fields

The five most commonly assessed medical fields in meta-analyses of test accuracy were oncology (25 percent), cardiovascular disease (21 percent), gastrointestinal disease (16 percent), obstetrics and gynecology (15 percent), and infectious disease (13 percent). For the comparison between fields we excluded 92 studies (16 percent of those relevant to the top 5 fields) that were considered relevant to more than one of these fields. There were few differences in the methods or reporting of meta-analyses in these fields (Tables 7a and 7b). There was a higher rate of exclusion of studies using quality criteria in meta-analyses in gastrointestinal (28 percent) and cardiovascular (25 percent) disease compared to those in obstetrics and gynecology (12 percent), infectious disease (13 percent), or oncology (17 percent); overall P = 0.022. Additionally, there was a higher rate of exclusion of studies based on sample size cut-offs in reviews in oncology (31 percent), cardiovascular disease (26 percent), and infectious disease (23 percent), compared to those in obstetrics and gynecology (9 percent) or gastrointestinal disease (17 percent); overall P = 0.003. Due to the large number of comparisons performed these differences may be chance findings.

Table 7a. Comparison of meta-analyses of test accuracy conducted in the five medical fields where most reviews had been published: number of studies, tests, and reference standards

Characteristics	Cardiovascular Disease (n=156)	Obstetrics & Gynecology (n=77)	Gastrointestinal Disease (n=65)	Infectious Disease (n=77)	Oncology (n=120)	P-value
Number of studies, median [25th–75th percentile]	20 [12–40]	16 [10–24]	17 [9–29]	19 [11–32]	18 [11–30]	0.099
Number of index tests, median [25th–75th percentile]	2 [1–3]	1 [1–2]	2 [1–3]	1 [1–2]	1 [1–2]	0.047
Number of reference standards, median [25th–75th percentile]	1 [1–1]	1 [1–2]	1 [1–1]	1 [1–1]	1 [1–1]	0.006

Table 7b. Comparison of meta-analyses of test accuracy conducted in the five medical fields where most reviews had been published

Characteristics	Cardiovascular Disease (n=156) n (%)	Obstetrics & Gynecology (n=77) n (%)	Gastrointestinal Disease (n=65) n (%)	Infectious Disease (n=77) n (%)	Oncology (n=120) n (%)	P-value
Searching and study selection: Exact search provided	46 (29)	17 (22)	17 (26)	19 (25)	25 (21)	0.540
Searching and study selection: Years searched	147 (94)	73 (95)	60 (92)	70 (91)	112 (93)	0.843
Searching and study selection: Exclusion based on quality criteria	39 (25)	9 (12)	18 (28)	10 (13)	20 (17)	0.022
Searching and study selection: Exclusion based on minimum sample size	40 (26)	7 (9)	11 (17)	18 (23)	37 (31)	0.003
Searching and study selection: MEDLINE plus at least one additional database	66 (42)	42 (55)	36 (55)	45 (58)	55 (46)	0.091
Quality assessment: Settings	63 (40)	21 (27)	23 (35)	34 (44)	43 (36)	0.220
Quality assessment: Consecutive patients	63 (40)	41 (53)	24 (37)	39 (51)	44 (37)	0.077
Quality assessment: Prospective patient sampling	80 (51)	54 (70)	39 (60)	45 (58)	70 (58)	0.103
Quality assessment: Any blinding	106 (68)	43 (56)	39 (60)	60 (78)	77 (64)	0.040
Quality assessment: Verification bias	70 (45)	38 (49)	29 (45)	37 (48)	60 (50)	0.902
Quality assessment: QUADAS	30 (19)	8 (10)	15 (23)	21 (27)	21 (18)	0.086
Quality assessment: STARD	13 (8)	7 (9)	6 (9)	7 (9)	11 (9)	0.996
Provides data for re-analysis	92 (59)	50 (65)	37 (57)	44 (57)	79 (66)	0.576

Table 7b. Comparison of meta-analyses of test accuracy conducted in the five medical fields where most reviews had been published (continued)

Characteristics	Cardiovascular Disease (n=156) n (%)	Obstetrics & Gynecology (n=77) n (%)	Gastrointestinal Disease (n=65) n (%)	Infectious Disease (n=77) n (%)	Oncology (n=120) n (%)	P-value
Heterogeneity testing	84 (54)	51 (66)	39 (60)	42 (55)	61 (51)	0.253
Random effects methods	96 (62)	44 (57)	39 (60)	39 (51)	61 (52)	0.385
Advanced methods	17 (11)	7 (9)	7 (11)	5 (6)	14 (12)	0.809

QUADAS = Quality Assessment of Diagnostic Accuracy Studies; STARD = Standards for Reporting of Diagnostic Accuracy
P-values are from Fisher exact tests for nominal variables and Kruskal-Wallis tests for continuous or count variables. Studies investigating tests belonging to different categories have been excluded. Meta-analyses that included index tests belonging to more than one category have been excluded from this analysis.

Comparison of Reviews of the Five Most Commonly Assessed Test Categories

The five most commonly assessed test categories were imaging tests (44 percent), biomarkers (28 percent), aspects of the clinical examination (15 percent), histological tests (14 percent, including cytological and culture-based tests), and physiologic/challenge (5 percent) tests. For the comparison between test types we excluded 75 studies (11 percent of those relevant to the top 5 test categories) that considered tests belonging to more than one test type (for example, reviews of multiple index tests belonging to different categories). There were several significant differences in the reporting and methods characteristics of meta-analyses assessing different test types (Tables 8a and 8b). The most striking of these differences pertained to the methods used by the reviews to appraise the quality of primary studies. Generally reviews of histological, cytological or culture based tests were less likely to assess quality items such as assessor blinding, verification bias, and prospective or consecutive patient recruitment. The use of the QUADAS instrument to guide quality assessment was also less common in reviews of histological, cytological or culture based tests. Similar patterns were observed for physiologic tests (although the number of available reviews was substantially smaller).

Table 8a. Comparison of meta-analyses of test accuracy conducted in the five test categories assessed in most meta-analyses: number of studies, tests, and reference standards

Characteristics	Histological Tests (n=68)	Clinical Examination(n= 81)	Imaging Tests (n=296)	Biomarkers (n=172)	Physiologic Tests* (n=22)	P-value
Number of studies, median [25th–75th percentile]	19 [11–28]	14 [10–24]	19 [12–33]	21 [11–35]	11 [8–27]	0.010
Number of index tests, median [25th–75th percentile]	1 [1–2]	2 [1–4]	1 [1–2]	1 [1–2]	1 [1–1]	0.001
Number of reference standards, median	1 [1–1]	1 [1–1]	1 [1–1]	1 [1–1]	1 [1–1]	0.869

* Including challenge tests.

Table 8b. Comparison of meta-analyses of test accuracy conducted in the five test categories assessed in most meta-analyses

Characteristics	Histological Tests* (n=68) n (%)	Clinical Examination (n=81) n (%)	Imaging Tests (n=296) n (%)	Biomarkers (n=172) n (%)	Physiologic Tests† (n=22) n (%)	P-value
Searching and study selection: Exact search provided,	14 (21)	31 (38)	71 (24)	42 (24)	5 (23)	0.096
Years searched, n (%)	58 (85)	77 (95)	274 (93)	160 (93)	19 (86)	0.161
Searching and study selection: Exclusion based on quality criteria, n (%)	9 (13)	17 (21)	53 (18)	39 (23)	1 (5)	0.171
Searching and study selection: Exclusion based on minimum sample size, n (%)	10 (15)	9 (11)	93 (31)	25 (15)	6 (27)	<0.001
Searching and study selection: MEDLINE plus at least one additional database, n (%)	26 (38)	45 (56)	145 (49)	96 (56)	7 (32)	0.039
Quality assessment: Settings, n (%)	14 (21)	55 (68)	79 (27)	83 (48)	11 (50)	<0.001
Quality assessment: Consecutive patients, n (%)	24 (35)	36 (44)	121 (41)	85 (49)	4 (18)	0.030
Quality assessment: Prospective patient sampling, n (%)	32 (47)	44 (54)	193 (65)	104 (60)	8 (36)	0.007
Quality assessment: Any blinding, n (%)	29 (43)	58 (72)	202 (68)	117 (68)	14 (64)	0.001
Verification bias, n (%)	22 (32)	40 (49)	152 (51)	93 (54)	5 (23)	0.003
QUADAS, n (%)	4 (6)	14 (17)	66 (22)	45 (26)	1 (5)	0.001
STARD, n (%)	2 (3)	3 (4)	27 (9)	23 (13)	1 (5)	0.034
Provides data for re-analysis	47 (69)	42 (52)	177 (60)	101 (59)	13 (59)	0.325
Heterogeneity testing	31 (46)	44 (54)	184 (62)	110 (64)	10 (458)	0.034
Random effects methods	33 (49)	49 (60)	168 (57)	109 (63)	9 (41)	0.112
Advanced methods	3 (4)	8 (10)	38 (13)	15 (9)	2 (9)	0.280

QUADAS = Quality Assessment of Diagnostic Accuracy Studies; STARD = Standards for Reporting of Diagnostic Accuracy

P-values are from Fisher exact tests for nominal variables and Kruskal-Wallis tests for continuous or count variables. Studies investigating tests belonging to different categories have been excluded. Meta-analyses covered more than one of the relevant clinical topics have been excluded from this analysis.

* Including cytological and culture-based tests.

† Including challenge tests.

Journal Impact Factor and Citation Count

In regression analyses with citation count (of each meta-analysis) as the dependent variable, after adjusting for the effect of journal impact factor and accounting for the number of years since manuscript publication, the following factors were predictive of a higher number of citations (with $P<0.001$): whether the study did not evaluate aspects of the clinical examination ("not clinical exam"); a larger number of included studies; whether the meta-analysis was relevant to adult medicine ("not pediatrics"); whether results were presented graphically; whether advanced meta-analysis methods were used; whether the topic was not related to orthopedics; and whether comparisons between alternative index tests were reported.

In analyses assessing the association between journal impact factor and reporting characteristics of systematic reviews of test accuracy, higher journal impact factor was associated with the following characteristics (with $P<0.001$): assessing the accuracy of aspects of the clinical examination; reporting quantitative analyses using likelihood ratio metrics; use of random effects methods; availability of the full search strategy upon request; assessment of blinding in the primary studies; whether the review assessed the enrollment of consecutive patients in the primary studies; and whether the reference lists of relevant review articles were perused as part of the search strategy.

When comparing reviews published in high impact factor general medical journals versus reviews published in other journals, few reporting differences were observed (with $P<0.001$): high-impact factor journals were more likely to publish meta-analyses assessing the accuracy of aspects of the clinical examination and using random effects models.

We note that the above analyses are exploratory in nature, and may—to a large extent—reflect journal editorial policies. Appendix Tables F1–F3 present additional information from the analyses of citation counts and journal impact factors.

Discussion

We performed a comprehensive review of 760 medical test accuracy meta-analyses published over the last 25 years. This work provides a “snapshot” of the available literature and an overview of longitudinal trends in methods and reporting, with the aim of identifying where future reviews could be improved. Meta-analyses of test accuracy are increasingly being pursued: in recent years approximately 100 such reports have been published annually. Overall, the available literature appears to have several limitations (Box 1): most reviews do not appraise important quality items, statistical analyses use methods that may be suboptimal for test accuracy and direct comparisons of index tests are scarce. Our findings regarding the limitations of existing systematic reviews of test accuracy generally agree with previously published, smaller-scale surveys of reviews of test accuracy. We have summarized some of these previous empirical investigations in Table 9. Generally, previous assessments of systematic reviews of diagnostic tests have assessed much smaller numbers of studies or have been limited to a single clinical topic (e.g., oncology[31]). Furthermore, with the exception of a report focusing on the statistical methods used for meta-analysis,[11] no previous overview has included an adequate number of studies spread over several years that would allow the exploration of trends over time.

Box 1. Common limitations of existing systematic reviews of test accuracy

Comparative effectiveness of medical tests
- Most systematic reviews were focused on a single index test.
- Most meta-analyses did not consider (direct or indirect) comparisons between alternative index tests.

Literature search and study identification
- Many systematic reviews relied on a relatively limited number of databases for the identification of potentially eligible studies.
- Search strategies were often not provided in detail.

Selection of studies
- Selection of studies for inclusion in meta-analyses was frequently based on quality criteria. It was often unclear whether these criteria were predetermined in a review protocol.

Data-extraction and qualitative synthesis
- Many reviews did not provide adequate summaries of the included studies. Settings of test use, the expected role of the test, study design characteristics, and demographics of participants, were often not reported.
- The counts needed to reconstruct the 2×2 tables of results used in each study were often not provided.

Assessment of study quality
- The assessment of study quality was often limited. Validated checklists, such as the QUADAS, were not universally used.
- Quality assessment methods were non-standardized and operational definitions for individual quality items were not always explicitly provided.

Meta-analysis
- Assessment of statistical heterogeneity was often not performed or was incompletely reported.
- Regression methods were infrequently used to explore between-study heterogeneity.
- A substantial number of published reviews did not use random effects meta-analysis models; thus, summary estimates may not be generalizable to future studies.
- Advanced meta-analysis methods that account for the bivariate nature of sensitivity and specificity were infrequently used, even in recent years.

Table 9. Summary of selected previously published overviews of systematic review of test accuracy

Characteristics	Irwig, 1994[32]	Whiting, 2005[33]	Dinnes, 2005[34]	Mallet, 2006[31]	Moher, 2007[35]	Willis, 2011[11]	Current Project
Number of included SRs	11	114	189 (133 used statistical synthesis methods)	89 (25 assessed in detail)	23 (diagnostic/prognostic SRs among 300 SRs identified)	236	760
Selection criteria	All inclusive; meta-analysis of test accuracy as primary focus.	All inclusive	All inclusive	Cancer diagnosis; included SRs regardless of the use of quantitative synthesis methods. Screening tests + tests for risk factors were excluded; computer decision tools were also excluded.	All inclusive	All inclusive; reviews had to have searched ≥2 databases, stated search terms and inclusion criteria, and used a statistical method to summarize test accuracy.	All inclusive; reviews had to have used a statistical method to summarize test accuracy. Excluded HTAs, Cochrane reviews and AHRQ EPC reports.
Years covered	Jan 1990 – Dec 1991	1995 – 2001	Up to 2002	1990 – 2003	November 2004	Up to 2008	1966 – 2009
Databases searched	MEDLINE, experts, bibliographies of retrieved papers	DARE	DARE (update of the search used in Whiting et al.[33])	MEDLINE, Embase, MEDION, Cancerlit, HTA, DARE, Cochrane Database of Systematic Reviews	MEDLINE	MEDLINE, Embase, CINAHL, Cochrane Library, PsychInfo, Global health, HMIC, AMED	MEDLINE, bibliographies of papers and relevant reviews
Items extracted	Literature review methods; data extraction and presentation; statistical analysis methods.	Quality assessment methods.	Systematic review methods; statistical analysis and reporting; trends over time in statistical method use.	Objectives and setting of the SRs; participant characteristics. In the 25 studies assessed in detail: quality assessment methods; whether meta-analysis was performed; reporting of results; availability of data for reanalysis.	Basic reporting characteristics of reviews (bibliometric features, outcomes considered, whether any quantitative method was used).	Statistical and quality assessment methods; settings of test use; trends over time in statistical method use.	Literature search and study selection methods; quality assessment methods; statistical analyses and reporting methods; trends over time in multiple aspects of the review process, statistical analysis and reporting; availability of data for re-analysis.

Table 9. Summary of selected previously published overviews of systematic review of test accuracy (continued)

Characteristics	Irwig, 1994[32]	Whiting, 2005[33]	Dinnes, 2005[34]	Mallet, 2006[31]	Moher, 2007[35]	Willis, 2011[11]	Current Project
Main findings	2 of 11 studies reported the complete search strategy; all studies analyzed sensitivity and specificity, 2 studies used the sROC method and 2 did not provide a summary estimate; 6 of 11 studies discussed variability in reference standards; 7 studies reported comparisons between 2 or more index tests.	49% of SRs had not conducted quality assessment; in most cases information on quality was incorporated in narrative synthesis; 13% of reviews used quality as an inclusion criterion.	70% of SRs used quantitative methods; 52% used MEDLINE as the only source; 69% performed quality assessment; median number of studies=18; 68% of SRs do not report tests for heterogeneity (58% of those using statistical analyses); naïve pooling has decreased over time.	75% of SRs stated inclusion criteria, 40% reported details of study design, 17% reported on the clinical setting, 17% reported on disease severity, 49% reported on tumor stage. Of the 25 reviews assessed in detail, 56% reported sensitivity, specificity, and sample sizes for individual studies. Of the 89 reviews, 61% attempted to formally synthesize results of the studies and 32% reported formal assessments of study quality.	No SRs were updates of previous reviews; harms were considered in 54% and costs in 35% of the reviews were this information was considered relevant; median number of included studies=39; quantitative synthesis was performed in 48% of SRs.	27% of SRs used advanced statistical methods (BREM or hsROC); between 2006 and 2008 QUADAS was used in 40% of the studies; imaging tests are the most commonly assessed test category; 80% of tests are normally used in specialist settings.	As detailed in this report.

Studies are listed chronologically, based on the dates covered by their searches.

AHRQ = Agency for Healthcare Research and Quality; AMED = Allied and Complementary Medicine database; BREM = bivariate random effects meta-analysis; CINAHL = Cumulative Index to Nursing and Allied Health Literature; DARE = Database of Reviews of Effects; EPC = Evidence-based Practice Center; HMIC = Health Management Information Consortium database; hsROC = hierarchical summary receiver operating characteristic meta-analysis method; HTAs = health technology assessments; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; SR = systematic review; sROC = summary receiver operating characteristic method

We also found that many aspects of the methods and reporting of systematic reviews of medical test accuracy have improved over time. Searching of multiple electronic databases without language restrictions has become more common; quality items such as verification bias, spectrum bias and blinding have been increasingly been considered in quality appraisal; and advanced statistical methods that simultaneously model sensitivity and specificity are beginning to be adopted. Empirical studies comparing the reporting of methodological quality items have documented an increase in the clarity of reporting of quality items in systematic reviews of therapeutic interventions[36] after the International Committee of Medical Journal Editors endorsed the Quality of Reporting of Meta-analyses (QUOROM) checklist, compared to before. Similar data exist on the impact of the Consolidated Standards of Reporting Trials (CONSORT) statement[37,38] for reporting of randomized trials.[39,40] We observed that the QUADAS tool[19,21] and quality items developed based on the STARD reporting checklist[22,23] were used more often by recent systematic reviews; we hypothesize that their use may have had a similar influence on the reporting of meta-analyses of test accuracy.

A recent focused empirical assessment[11] of meta-analyses of medical tests concluded that the increased use of bivariate random effects statistical models for sensitivity and specificity coincided with the development of easy-to-use routines for performing such analyses (e.g., the metandi and midas commands in Stata or scripts for SAS programming).[41,42] We observed the same pattern. Further, we observed that, at the same time, the use of simpler, but less appropriate methods such as the fixed effects SROC model of Moses and Littenberg[13,27] (which accounts for only part of the uncertainty in the bivariate probability model), has decreased. Guidance within the AHRQ Evidence-based Practice Center Program[c] and the Cochrane Collaboration[d] supports the use of hierarchical modeling methods in meta-analyses of test accuracy. Although such models are more theoretically motivated compared to separate univariate analyses of sensitivity and specificity for the binary classification case, their judicious application requires an understanding of the underlying model assumptions.[43]

We found substantial differences in methods and reporting of test accuracy studies across different types of medical tests. These differences may reflect either heterogeneous diffusion of methodological advances between research groups focusing on specific test types, or the reviewers' assessment that specific methodological approaches are not applicable to specific test types. In contrast, we found few differences in comparisons across different medical fields. Differences in some reporting or methodological characteristics of reviews correlated with the impact factor of journals where they were published and the number of citations they accrued over time. Interpretation of these differences is challenging, given the large number of comparisons performed and the possibility that journal editorial policies and journal readability could confound many of the observed associations.

Our work has several limitations that need to be considered when interpreting our results. First, we relied on searches using methodological filters for identifying reports of meta-analyses of medical test accuracy studies and we only considered English language publications.[44] Also, we relied exclusively on MEDLINE searches, supplemented by screening of the reference lists of eligible studies and those of relevant review articles, to identify eligible reviews. More comprehensive searches would have required the examination of a much larger number of

[c] Available at: www.effectivehealthcare.ahrq.gov/tasks/sites/ehc/assets/File/methods_guide_for_medical_tests.pdf; accessed December 27, 2011.
[d] Available at: http://srdta.cochrane.org/handbook-dta-reviews; accessed December 27[th], 2011.

abstracts with little expected incremental yield. Second, we focused on meta-analyses using aggregate published data and excluded individual patient data analyses. Although the latter represent a minority of all published meta-analyses they provide additional flexibility in exploring between-study heterogeneity due to patient level factors.[45] Third, we did not perform double extraction for all eligible studies. However, we implemented several procedures for standardizing the definition of the extracted variables during data extraction and performed extensive quality control of the final dataset. Further, a substantial proportion of eligible articles were extracted in duplicate.

This comprehensive overview of meta-analyses of test accuracy highlights the current status and the temporal evolution of a complex research field. Available meta-analyses of medical tests have several limitations in regards to methodological approaches and reporting characteristics; however, over time reviews have increasingly performed more comprehensive assessments of study quality and have used more appropriate statistical methods addressing the particular challenges relevant to reviews of test accuracy. Based on our review of the literature, and observations from this current empirical assessment, we identify some cross-cutting methodological issues relevant to meta-analytic practice in Box 2. Areas for potential future methodological research include the assessment of publication and reporting bias in reviews of test accuracy, the collection of empirical evidence on how study-level characteristics can influence the results of systematic reviews, quantitative methods for the comparative assessment of multiple alternative index tests, and the evaluation of modeling approaches for contextualizing the findings of reviews of test accuracy.

The large and rapidly expanding number of available meta-analyses identified by this overview reflects the growing interest in "evidence-based diagnosis".[46,47] Increasing use of quality checklists is expected to facilitate further improvements in the quality assessment of primary studies included in meta-analyses of medical tests. Similarly, increasing diffusion of methodological advances, availability of software to perform advanced statistical analyses and clear guidelines for the conduct and reporting of meta-analyses of test accuracy will hopefully lead to further improvements in the practice of systematic reviews of medical tests.

Box 2. Cross-cutting methodological issues relevant to meta-analytic practice

Comparative effectiveness reviews of medical tests[48]

- Often many alternative tests are applicable to a given testing scenario; systematic reviews may want to consider all relevant index tests that are applicable to the population and disease of interest.
- In cases where multiple index tests are applicable, reviews that directly compare test performance may have the greatest impact on clinical practice.

Defining the setting and role of test use

- The findings of systematic reviews can be meaningfully applied to clinical practice only if the role (add-on, triage, replacement) and setting (screening, diagnosis, prognosis/prediction, treatment selection) of test use is explicitly considered. These aspects of test use have implications for the study designs to be considered, the information to be extracted from each eligible study, the interpretation of individual study results, and the synthesis of findings across studies.[49]

Box 2. Cross-cutting methodological issues relevant to meta-analytic practice (continued)

Methods and reporting of systematic reviews of test performance[50-53]

- *A priori* defined protocols, clearly delineating the scope of the review and outlining the proposed methods is in accordance with commonly held standards of research conduct, and probably applies to systematic reviews as much as other research enterprises.
- Explicitly reporting the methodological approach followed by systematic reviews (including any deviations from the review protocol) promotes clarity. It is probably good practice for reviewers to consult existing (and continuously evolving) guidance on the optimal methods for searching the literature, identifying and selecting relevant studies, extracting data, assessing the validity of included studies, qualitatively and quantitatively synthesizing study results.
- The assessment of study "quality" or "risk of bias" is an important component of systematic reviews of medical tests – yet exactly how these assessments (should) affect the conclusions of the systematic review is still a matter of research.

Examining the applicability of review findings and transferability of estimates

- Based on our empirical assessment, many reviews did not adequately describe whether the applicability of research findings from individual studies was assessed, and if yes, how. Information for the assessment of applicability includes, but is not limited to, details about the index and reference standard tests used, and the demographics and disease-related characteristics of the population enrolled in each study. Assessment of applicability is critical in contextualizing the conclusions of a systematic review.

Interpreting and contextualizing review results

- Reviews of test accuracy address an intermediate component of the effect of tests on clinical outcomes.[54] Although studies assessing the overarching question of test effectiveness on clinical outcomes are rare, reviews of test performance often can only provide part of the information needed to fully assess the impact of tests.[55] Other intermediate outcomes that could be considered include the impact of test results on physicians' diagnostic thinking, and on therapeutic decisionmaking.[56]
- In the absence of studies assessing the direct effect of tests on clinical outcomes, formal modeling or simple ("back-of-the-envelope") projections of the potential impact of tests may be informative.[57]

References

1. Miettinen OS. The modern scientific physician: 3. Scientific diagnosis. CMAJ. 2001;165:781-82.
2. Lijmer JG, Leeflang M, Bossuyt PM. Proposals for a phased evaluation of medical tests. Med Decis Making. 2009;29:E13-E21.
3. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? Ann Intern Med. 2006;144:850-5.
4. Hayen A, Macaskill P, Irwig L, et al. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. J Clin Epidemiol. 2010;63:883-91.
5. Gatsonis C, Paliwal P. Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer. AJR Am J Roentgenol. 2006;187:271-81.
6. Reitsma JB, Glas AS, Rutjes AW, et al. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol. 2005;58:982-90.
7. van Houwelingen HC, Zwinderman KH, Stijnen T. A bivariate approach to meta-analysis. Stat Med. 1993;12:2273-84.
8. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. Stat Med. 2002;21:589-624.
9. Rutter CM, Gatsonis CA. Regression methods for meta-analysis of diagnostic test data. Acad Radiol. 1995;2 Suppl 1:S48-S56.
10. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med. 2001;20:2865-84.
11. 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.
12. Glas AS, Lijmer JG, Prins MH, et al. The diagnostic odds ratio: a single indicator of test performance. J Clin Epidemiol. 2003;56:1129-35.
13. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med. 1993;12:1293-316.
14. 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.
15. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med. 1978;299:926-30.
16. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA. 1999;282:1061-6.
17. Rutjes AW, Reitsma JB, Di NM, et al. Evidence of bias and variation in diagnostic accuracy studies. CMAJ. 2006;174:469-76.
18. Westwood ME, Whiting PF, Kleijnen J. How does study quality affect the results of a diagnostic meta-analysis? BMC Med Res Methodol. 2005;5:20.
19. Whiting P, Rutjes AW, Reitsma JB, et al. 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.
20. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. Health Technol Assess. 2004;8:iii, 1-iii234.
21. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006;6:9.
22. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Ann Intern Med. 2003;138:40-4.

23. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med. 2003;138:W1-12.

24. Irwig L, Macaskill P, Glasziou P, et al. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol. 1995;48:119-30.

25. Steingart KR, Henry M, Laal S, et al. A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Postgrad Med J. 2007;83:705-12.

26. Steingart KR, Henry M, Laal S, et al. A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Thorax. 2007;62:911-8.

27. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making. 1993;13:313-21.

28. Chu H, Cole SR. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol. 2006;59:1331-2.

29. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics. 2007;8:239-51.

30. Arends LR, Hamza TH, van Houwelingen JC, et a;. Bivariate random effects meta-analysis of ROC curves. Med Decis Making. 2008;28:621-38.

31. Mallett S, Deeks JJ, Halligan S, et al. Systematic reviews of diagnostic tests in cancer: review of methods and reporting. BMJ. 2006;333:413.

32. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-76.

33. Whiting P, Rutjes AW, Dinnes J, et al. A systematic review finds that diagnostic reviews fail to incorporate quality despite available tools. J Clin Epidemiol. 2005;58:1-12.

34. Dinnes J, Deeks J, Kirby J, et al. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. Health Technol Assess. 2005;9:1-113, iii.

35. Moher D, Tetzlaff J, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews. PLoS Med. 2007;4:e78.

36. Wen J, Ren Y, Wang L, et al. The reporting quality of meta-analyses improves: a random sampling study. J Clin Epidemiol. 2008;61:770-5.

37. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. Ann Intern Med. 2001;134:657-62.

38. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann Intern Med. 2001;134:663-94.

39. Moher D, Jones A, Lepage L. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. JAMA. 2001;285:1992-5.

40. Plint AC, Moher D, Morrison A, et al. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust. 2006;185:263-7.

41. Harbord RM. metandi: Meta-analysis of diagnostic accuracy using hierarchical logistic regression. In: Sterne J, ed. Meta-analysis: an updated collection from the Stata Journal. Stata Press; 2011.

42. Macaskill P. Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis. J Clin Epidemiol. 2004;57:925-32.

43. Chappell FM, Raab GM, Wardlaw JM. When are summary ROC curves appropriate for diagnostic meta-analyses? Stat Med. 2009;28:2653-68.

44. Wilczynski NL, Haynes RB. Consistency and accuracy of indexing systematic review articles and meta-analyses in medline. Health Info Libr J. 2009;26:203-10.

45. Berlin JA, Santanna J, Schmid CH, et al. Individual patient- versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. Stat Med. 2002;21:371-87.

46. Fowler PB. Evidence-based diagnosis. J Eval Clin Pract. 1997;3:153-9.

47. Whiting P, Harbord R, de Salis, I, et al. Evidence-based diagnosis. J Health Serv Res Policy. 2008;13 Suppl 3:57-63.

48. Pandharipande PV, Gazelle GS. Comparative effectiveness research: what it means for radiology. Radiology. 2009;253:600-5.

49. Hayen A, Macaskill P, Irwig L, et al. Appropriate statistical methods are required to assess diagnostic tests for replacement, add-on, and triage. J Clin Epidemiol. 2010;63:883-91.

50. Tatsioni A, Zarin DA, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies. Ann Intern Med. 2005;142:1048-55.

51. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med. 1994;120:667-76.

52. Deville WL, Buntinx F, Bouter LM, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. BMC Med Res Methodol. 2002;2:9.

53. Deeks JJ. Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests. BMJ. 2001;323:157-62.

54. Lijmer JG, Bossuyt PM. Various randomized designs can be used to evaluate medical tests. J Clin Epidemiol. 2009;62:364-73.

55. Lord SJ, Irwig L, Simes RJ. When is measuring sensitivity and specificity sufficient to evaluate a diagnostic test, and when do we need randomized trials? Ann Intern Med. 2006;144:850-5.

56. Fineberg HV, Bauman R, Sosman M. Computerized cranial tomography. Effect on diagnostic and therapeutic plans. JAMA. 1977;238:224-7.

57. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. Med Decis Making. 2009;29:E22-9.

Abbreviations

ANOVA	analysis of variance
AHRQ	Agency for Healthcare Research and Quality
CI	confidence interval
EPC	Evidence-based Practice Center
OR	odds ratio
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
QUOROM	Quality of Reporting of Meta-analyses
ROC	receiver operating characteristic
STARD	STAndards for the Reporting of Diagnostic Accuracy Studies

Appendix A. Search Strategy

```
Ovid MEDLINE (1966-2009)
1. exp "sensitivity and specificity"/
2. exp Predictive Value of Tests/
3. exp ROC CURVE/
4. exp Mass Screening/
5. exp diagnosis/
6. exp REPRODUCIBILITY OF RESULTS/
7. exp false negative reactions/ or false positive reactions.mp.
8. predictive value.tw.
9. (sensitivity or specificity).tw.
10. accuracy.tw.
11. screening.tw.
12. roc.tw.
13. reproducibility.tw.
14. (false positive or false negative).tw.
15. likelihood ratio.tw.
16. accuracy.tw.
17. di.fs.
18. or/1-17
19. limit 18 to meta analysis
20. systematic review$.tw.
21. meta analy$.tw.
22. 20 or 21
23. exp Meta-Analysis/
24. 18 and 23
25. 18 and 22
26. 19 or 24 or 25
27. limit 26 to human
28. limit 27 to english language
```

Appendix B. Data Extraction Form

```
ui                  ##############
refid               ##########
extractor           _______________________
first.author        _______________________
title  ________________________________________________________________________
journal ________________________________________________________________________
pub.year            ####

**********************************************************************************

[Eligibility criteria:]
[1. Meta-analysis of diagnostic or prognostic accuracy ]
[2. Was based on systematic review methodology (key question; search; elgibility
criteria)]
[3. Has an imaging, clinical, genetic test - (not risk instruments like APACHE)]
[4. Exclude comparison of continuous measurements ]
[NB the above exclude reviews of clinical outcomes after Dx test application]

[1 = yes; 0 = no]

eligible            #           [Is this study eligible?  ]
excl.reason
________________________________________________________________________________


**********************************************************************************

/* This is for the performed analyses */

n.studies            ###      [Total articles in paper - e.g. number at the end of the
                                    flowchart, N of papers in metanalysis]

n.tests              ##       [Number of different index tests included in any meta-
                                    analysis: e.g.,  CT or MRI or PET vs. a gold
                                    standard]
n.ref.stand          ##       [Number of different reference tests included in any meta-
                                    analysis: e.g.,  outcomes - breast Ca, lung Ca etc ]

[Another way to look at it: (n.tests) x (n.ref.stand) is giving us the number of meta-
analyses in the paper]

**********************************************************************************

[Try to classify the test that is being studied (medical field)]
field.cvd           #        [cardiovascular]
field.obgyn         #        [obstetrics-gynecology]
field.gi            #        [gastrointestinal]
field.id            #        [infectious diseases]
field.onc           #        [hematology-oncology]
field.kidn          #        [nephrology - urology]
field.rheu          #        [rheumatology]
field.pulm          #        [pulmonary medicine]
field.orth          #        [orthopedics]
field.psych         #        [psychiatry]
field.ent           #        [Ear - Nose - Throat]
field.neuro         #        [Neurology]
field.peds          #        [Pediatrics]


field.other         ________________________________________________________________
[if it does not fit to the above list,  free text]


**********************************************************************************
[write 1-Y or 0-N to the following questions]
binary.test         #        [is this a binary test? - based on
                                    the handling of studies in the meta-analysis ]

**********************************************************************************

[Try to categorise the type of test that the paper studies]

histology.cytology    #        [e.g., biopsy, pap smear]
clinical.exam         #        [e.g., signs (mcburney's, boas), murmurs,
                                    blood pressure measurement]
```

```
imaging              #        [e.g., CT scan, V-Q scan, MRI, chest X-ray]
biomarker            #        [e.g., PSA, BNP, PTH]
clinical.test        #        [e.g., this can be a challenge test such as using levodopa
                                    for idiopathic Parkinson's diagnosis, or doing a
                                    treadmill test (stress test) or doing a tilt test]
physiologic.test     #        [e.g., measuring electric impendance, EKG, EEG, doppler
                                    measurement of blood flow (not image, but Qa), sleep
                                    apnea testing, spirometry]
endoscopy.exams      #        [e.g., gastroscopy, colonoscopy]

[If there are no obvious candidates for the above, please describe in 80 chars]
describe.test       ______________________________________________________________


**********************************************************************************
[Describe the search - again 1 = yes; 0 = no]

exact.search.desc   #    [Do they report the *exact* search in a
                                    way that can be replicated?]

[If the above is 0 then do they report the following]
search.terms        #    [they list or enumerate search terms without boolean
                                    operands]
search.on.demand    #    [they state that the exact search is available upon request,
                                    available on a website, or in a previous paper]

yr.searched         #    [search years mentioned]
qual.exc            #    [in the inclusion criteria do they describe excluding studies
                                     based on study quality
                                    We do count exclusions based on
                                    - risk of verification bias
                                    - timing between index and reference test
                                           administration
                                    - blinding etc.
                                    We do not count study design characteristics
                                    -   i.e., exclusion of retrospective or Xsectional
                                           studies ]
min.n               #    [in the inclusion criteria do they describe excluding studies
                                    based on min sample size?]
min.no.sub          #### [if yes above, fill in cutoff sample size]

Which languages were included? [1] English only --- [2] English + specific other
--- [3] no restrictions or all --- [4] not stated
language            #

Which of the following databases were searched?
Medline             #   [MEDLINE - ANY VENDOR, Pubmed or OVID]
Embase              #   [EMBASE]
Conference          #   [conference proceedings]
biblio              #   [reference lists]
review              #   [review papers]
sci                 #   [Science citation index]
cc                  #   [current contents]
experts             #   [Contact experts in the field]
manufacturers       #   [were manufacturers specifically asked for info on studies]
specific.db         #   [Other specific database]
manual              #   [handsearching]
cochrane            #   [Cochrane CENTRAL or Cochrane database of systematic reviews]
CINAHL              #   [CINAHL]
unpublished.data    #   [Did they search for unpublished data]

Which funder?  [0] No funder  [1] Non-industry only --- [2] Any industry funding ---  [3]
not mentioned
funding             #

**********************************************************************************

[Did the meta-analysis abstract for each individual study the following characteristics?]
[NB -Answer yes if they describe abstracting this information even if they
       do not report it in a table, or even if they do not perform or report analyses by
       these characteristics -- answer 0 [no] if they do not explicitly state that they
       assessed the characteristic. Only note characteristics that were explicitly
       mentioned. ]

any.qual            #   [did they do any quality assessment - scores or items both
count]
settings            #   [setting of study -- e.g., tertiary care, rural]
consecutive         #   [whether participants were recruited consecutively or not]
prospective         #   [whether participants were recruited prospectively or not]
refstd              #   [the exact definition of the reference standard per study]
reader              #   [whether the test reader/assessor in each study was
```

```
                          experienced or not]
blinding.index       #   [blinding of index test assessor to reference standard
                          or to clinical information ]
blinding.refst       #   [blinding of reference test assessor to index test results]
blinding.unspecif    #   [they mention blinding but not distinguish in the above]
Age                  #   [describe age distribution in studies e.g., mean, sd]
Gender               #   [male/female]
Location             #   [geographic location e.g. US/Europe]
Spectrum bias        #   [representativeness of spectrum of patients studied or
                              description of severity]
Selection bias       #   [what criteria were used to select patients for study]
Time                 #   [adequacy of the time interval between the index test and
                              reference standard]
Test Independence    #   [was the reference standard independent of the index test?
                              i.e. the index test did not form part of the reference
                              standard]
Indeterminant results#   [were uninterpretable/intermediate test results reported?]
Withdrawals          #   [were withdrawals from the study explained?]
verbias              #   [verification bias - whether the decision to apply the reference
                              test is influenced by the results of the index test]

[If they  used QUADAS to rate studies indicate 1 below.  If not, indicate 0. ]

Quadas               #
Stard                #

[ IMPORTANT NOTE for HANDLING QUALITY EXTRACTIONS:
When the authors claim to have used QUADAS or STARD
   First, check the corresponding checklist (QUADAS or STARD) above.
   Then,
        - If they report the specific items they used
            (which may be a subset of the checklist) then check ONLY the items they used.
        - If they do not report any specific items from that checklist
             check ALL those that correspond to checklist items. ]
*************************************************************************************

With respect to the index test classification as (+) or (-) did the analysis
examine only a single threshold or did it examine multiple thresholds (analysed in any
way?)
Indicate
1= single threshold
2= multiple, analysed in separate meta-analyses
3= multiple, analysed in a single model (ordinal)

n.thresholds         #    [based on the handling of the test in the meta-analysis]

Does the paper provide data to repeat analyses?
has.counts           #    [Yes, if they report counts for the 2x2 tables, i.e., TP, FN,
                              FP, TN]
can.calculate.counts #    [If the above is no, do they provide sufficient statistics to
                              calculate counts e.g., sensitivity and N diseased, with
                              specificity and N nondiseased OR sensitivity, specificity,
                              prevalence and overall sample size OR sensitivity,
                              specificity and their CI's ]

*************************************************************************************
ANALYSES - GENERAL

rem.used             #    [Yes if they used any random effects model in their analyses]
bayes.used           #    [Yes if they used any bayesian approach in their analyses
                              using just Bayes rule does not count
                              as a Bayesian analysis]

*************************************************************************************
ANALYSES - METRICS

Which of the following metrics were reported or anyhow analysed/calculated?
Please check only the metrics that were used in a synthesis or to interpret a synthesis.
For example: A meta-analysis using the bivariate method synthesizes Sensitivity and
Specificity
(should be checked). If it then takes the summary Sensitivity and Specificity and
calculates
summary LR+ and LR- to aid in interpretation, then we should check LR also.

or                   #   [ diagnostic OR ]
sens                 #   [ sensitivity ]
spec                 #   [ specificity ]
LR                   #   [ likelihood ratios ]
Acc                  #   [ accuracy ]
pv                   #   [ predictive values ]
Q.star               #   [ Synthesis of Q* - i.e. analysis based on Q* from primary
```

```
                                studies]
auc                   #   [ Synthesis of individual ROC AUC's - i.e.
                                analysis based on AUCs from primary studies]

***********************************************************************************
ANALYSES - GRAPHICS

plot.shown            #   [ Do they show any graphs plots for synthesis ]

Specify the type of graphs that are shown

forest.plot           #   [ Forest plot ]
roc.space.plot        #   [ Plot in the ROC space - sensitivity vs (1-) specificity ]


Other plot(s) related to Dx test analyses
plot.describe          ______________________________________________________________

***********************************************************************************
ANALYSES - HETEROGENEITY
Testing for heterogeneity

hetero.test           #    [ Did they do any test for heterogeneity or for differential
                            model fit between fixed and random effects models? ]

Exploring heterogeneity
0 = no exploration of heterogeneity, (or no heterogeneity to explore)
1 = subgroup analyses - excluding a single study
2 = exploration of heterogeneity with regression models

hetero.explore        #    [ for studies that do both meta-regression and subgroup
                                 analyses enter "2" ]

***********************************************************************************
ANALYSES - MODELS

Do they perform univariate analyses? E.g., separate analyses of sensitivity,
specificity, analysis of ORs, AUCs, LRs and so on. Note that SROC analyses should not
be logged here - they should be logged under SROC/HSROC analyses

univariate            #    [ analyses done one outcome-at-a-time ]
naive.univariate      #    [ Do they do naive "pooling"? Examples are summing up numerator
                                for sensitivity and specificity, or weighting by size
                                or getting an unweighted mean ]


Do they perform advanced multivariate analyses? E.g., bivariate model (joint analysis of
 sensitivity and specificity) or the HSROC model (joint analysis of alpha and theta).
 Note that we do not record meta-regressions with  multiple predictors here.
0 = no advanced analyses
1 = bivariate model
2 = HSROC model
3 = bivariate and HSROC models  (e.g., if they show the summary point and the line)
4 = other (e.g. multiple thresholds, or a custom model that is complex) - free text

advanced              #
other.advanced        __________________________________________________


If they perform SROC/HSROC analyses, what method do they use?
1 = Moses and Littenberg
2 = Rutter and Gatsonis
3 = other (e.g., random intercept variation of Moses, or major axis regression) - free
               text

sroc.model            #
sroc.other            __________________________________________________

***********************************************************************************
ANALYSES - COMPARATIVE

Do they perform formal comparative analyses between 2 or more index tests, based on
 statistical procedures? Qualitative comparisons (eyeball, or based on overlap of CIs),
in the absence of a formal statistical test DO NsOT COUNT as comparative.

comparative          #

If yes, where the comparisons direct or indirect?
Direct:  (1) Both index tests were given to the same patients in each study and were
                      assessed against the same reference standard
         (2) This design was taken into account in a hierarchical model
```

```
Indirect: All that do not fall under the above

direct.comparisons  #

What statistical tests were used for comparing the performance characteristics of index
tests?
z.score             #   [ comparisons of summary estimates ]
meta.regression     #   [ test is a covariate in the HSROC or bivariate or other model ]
State something about the comparison method:
compare.method      ________________________________________________________________

************************************************************************************
```

Appendix C. List of Included Studies

1. Abbas SM, Bissett IP, Parry BR**.** Meta-analysis of oral water-soluble contrast agent in the management of adhesive small bowel obstruction. Br J Surg. 2007;94(4):404-11.
2. Abdelmoneim SS, Dhoble A, Bernier M, et al. Quantitative myocardial contrast echocardiography during pharmacological stress for diagnosis of coronary artery disease: a systematic review and meta-analysis of diagnostic accuracy studies. Eur J Echocardiogr. 2009;10(7):813-25.
3. Abdulla J, Abildstrom SZ, Gotzsche O, et al**.** 64-multislice detector computed tomography coronary angiography as potential alternative to conventional coronary angiography: a systematic review and meta-analysis. Eur Heart J. 2007;28(24):3042-50.
4. Abdulla J, Sivertsen J, Kofoed KF, et al. Evaluation of aortic valve stenosis by cardiac multislice computed tomography compared with echocardiography: a systematic review and meta-analysis. J Heart Valve Dis. 2009;18(6):634-43.
5. Abulafia O, Sherer DM**.** Automated cervical cytology: meta-analyses of the performance of the AutoPap 300 QC System. Obstet Gynecol Surv. 1999;54(7):469-76.
6. Adams K, Shah PL, Edmonds L, et al**.** Test performance of endobronchial ultrasound and transbronchial needle aspiration biopsy for mediastinal staging in patients with lung cancer: systematic review and meta-analysis. Thorax. 2009;64(9):757-62.
7. Aertgeerts B, Buntinx F, Kester A**.** The value of the CAGE in screening for alcohol abuse and alcohol dependence in general clinical populations: a diagnostic meta-analysis. J Clin Epidemiol. 2004;57(1):30-9.
8. Ahmad S, Beckett MW**.** Value of serum prolactin in the management of syncope. Emerg Med J. 2004;21(2):e3.
9. Akcil M, Karaagaoglu E, Demirhan B**.** Diagnostic accuracy of fine-needle aspiration cytology of palpable breast masses: an SROC curve with fixed and random effects linear meta-regression models. Diagn Cytopathol. 2008;36(5):303-10.
10. Alongi F, Ragusa P, Montemaggi P, et al**.** Combining independent studies of diagnostic fluorodeoxyglucose positron-emission tomography and computed tomography in mediastinal lymph node staging for non-small cell lung cancer. Tumori. 2006;92(4):327-33.
11. Alvarez Amezaga J, Barbier Herrero L, Pijoan del Barrio JI, et al. Diagnostic efficacy of sentinel node biopsy in oral squamous cell carcinoma. Cohort study and meta-analysis. Med Oral Patol Oral Cir Bucal. 2007;12(3):E235-43.
12. Alvarez S, Anorbe E, Alcorta P, et al**.** Role of sonography in the diagnosis of axillary lymph node metastases in breast cancer: a systematic review. AJR Am J Roentgenol. 2006;186(5):1342-8.
13. Anderson BA, Salem L, Flum DR**.** A systematic review of whether oral contrast is necessary for the computed tomography diagnosis of appendicitis in adults. Am J Surg. 2005;190(3):474-8.
14. Andersson RE**.** Meta-analysis of the clinical and laboratory diagnosis of appendicitis. Br J Surg. 2004;91(1):28-37.
15. Annovazzi A, Bagni B, Burroni L, et al**.** Nuclear medicine imaging of inflammatory/infective disorders of the abdomen. Nucl Med Commun. 2005;26(7):657-64.
16. Arbyn M, Buntinx F, Van Ranst M, et al**.** Virologic versus cytologic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst. 2004;96(4):280-93.

17. Arbyn M, Martin-Hirsch P, Buntinx F, et al**.** Triage of women with equivocal or low-grade cervical cytology results: a meta-analysis of the HPV test positivity rate. J Cell Mol Med. 2009;13(4):648-59.
18. Arbyn M, Paraskevaidis E, Martin-Hirsch P, et al**.** Clinical utility of HPV-DNA detection: triage of minor cervical lesions, follow-up of women treated for high-grade CIN: an update of pooled evidence. Gynecol Oncol. 2005;99(3 Suppl 1):S7-11.
19. Arbyn M, Schenck U**.** Detection of false negative Pap smears by rapid reviewing. A metaanalysis. Acta Cytol. 2000;44(6):949-57.
20. Arbyn M, Schenck U, Ellison E, et al**.** Metaanalysis of the accuracy of rapid prescreening relative to full screening of pap smears. Cancer. 2003;99(1):9-16.
21. Ashoke R, Brown LC, Rodway A, et al. Color duplex ultrasonography is insensitive for the detection of endoleak after aortic endografting: a systematic review. J Endovasc Ther. 2005;12(3):297-305.
22. Atieh MA**.** Accuracy of real-time polymerase chain reaction versus anaerobic culture in detection of Aggregatibacter actinomycetemcomitans and Porphyromonas gingivalis: a meta-analysis. J Periodontol. 2008;79(9):1620-9.
23. Attia J, Hatala R, Cook DJ, et al**.** The rational clinical examination. Does this adult patient have acute meningitis? JAMA. 1999;282(2):175-81.
24. Bachmann LM, Haberzeth S, Steurer J, et al**.** The accuracy of the Ottawa knee rule to rule out knee fractures: a systematic review. Ann Intern Med. 2004;140(2):121-4.
25. Bachmann LM, Kolb E, Koller MT, et al**.** Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. BMJ. 2003;326(7386):417.
26. Bachmann MO, Nelson SJ**.** Impact of diabetic retinopathy screening on a British district population: case detection and blindness prevention in an evidence-based model. J Epidemiol Community Health. 1998;52(1):45-52.
27. Badgett RG, Lucey CR, Mulrow CD**.** Can the clinical examination diagnose left-sided heart failure in adults? JAMA. 1997;277(21):1712-9.
28. Badgett RG, Mulrow CD, Otto PM, et al**.** How well can the chest radiograph diagnose left ventricular dysfunction? J Gen Intern Med. 1996;11(10):625-34.
29. Bafounta ML, Beauchet A, Aegerter P, et al**.** Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol. 2001;137(10):1343-50.
30. Bafounta ML, Beauchet A, Chagnon S, et al**.** Ultrasonography or palpation for detection of melanoma nodal invasion: a meta-analysis. Lancet Oncol. 2004;5(11):673-80.
31. Bagai A, Thavendiranathan P, Detsky AS**.** Does this patient have hearing impairment? JAMA. 2006;295(4):416-28.
32. Bailey B, Buckley NA, Amre DK**.** A meta-analysis of prognostic indicators to predict seizures, arrhythmias or death after tricyclic antidepressant overdose. J Toxicol Clin Toxicol. 2004;42(6):877-88.
33. Bailey JJ, Berson AS, Handelsman H, et al**.** Utility of current risk stratification tests for predicting major arrhythmic events after myocardial infarction. J Am Coll Cardiol. 2001;38(7):1902-11.
34. Baker PA, Depuydt A, Thompson JM**.** Thyromental distance measurement—fingers don't rule. Anaesthesia. 2009;64(8):878-82.
35. Bakis S, Irwig L, Wood G, et al**.** Exfoliative cytology as a diagnostic test for basal cell carcinoma: a meta-analysis. Br J Dermatol. 2004;150(5):829-36.
36. Balk EM, Ioannidis JP, Salem D, et al**.** Accuracy of biomarkers to diagnose acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med. 2001;37(5):478-94.

37. Banal F, Dougados M, Combescure C, et al**.** Sensitivity and specificity of the American College of Rheumatology 1987 criteria for the diagnosis of rheumatoid arthritis according to disease duration: a systematic literature review and meta-analysis. Ann Rheum Dis. 2009;68(7):1184-91.
38. Barnes CJ, Pietrobon R, Higgins LD**.** Does the pulse examination in patients with traumatic knee dislocation predict a surgical arterial injury? A meta-analysis. J Trauma. 2002;53(6):1109-14.
39. Basaran A, Basaran M**.** Diagnosis of acute appendicitis during pregnancy: a systematic review. Obstet Gynecol Surv. 2009;64(7):481-8; quiz 99.
40. Bastiaannet E, Groen H, Jager PL, et al. The value of FDG-PET in the detection, grading and response to therapy of soft tissue and bone sarcomas; a systematic review and meta-analysis. Cancer Treat Rev. 2004;30(1):83-101.
41. Bastian LA, Nanda K, Hasselblad V, et al**.** Diagnostic efficiency of home pregnancy test kits. A meta-analysis. Arch Fam Med. 1998;7(5):465-9.
42. Battaglia M, Pewsner D, Juni P, et al**.** Accuracy of B-type natriuretic peptide tests to exclude congestive heart failure: systematic review of test accuracy studies. Arch Intern Med. 2006;166(10):1073-80.
43. Bax JJ, Poldermans D, Elhendy A, et al**.** Sensitivity, specificity, and predictive accuracies of various noninvasive techniques for detecting hibernating myocardium. Curr Probl Cardiol. 2001;26(2):147-86.
44. Bax JJ, Wijns W, Cornel JH, et ak**.** Accuracy of currently available techniques for prediction of functional recovery after revascularization in patients with left ventricular dysfunction due to chronic coronary artery disease: comparison of pooled data. J Am Coll Cardiol. 1997;30(6):1451-60.
45. Beach J, Russell K, Blitz S, et al. A systematic review of the diagnosis of occupational asthma. Chest. 2007;131(2):569-78.
46. Beattie WS, Abdelnaem E, Wijeysundera DN, et al**.** A meta-analytic comparison of preoperative stress echocardiography and nuclear scintigraphy imaging. Anesth Analg. 2006;102(1):8-16.
47. Benatar M**.** A systematic review of diagnostic studies in myasthenia gravis. Neuromuscul Disord. 2006;16(7):459-67.
48. Benjaminse A, Gokeler A, van der Schans CP**.** Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. J Orthop Sports Phys Ther. 2006;36(5):267-88.
49. Benninger MS, Payne SC, Ferguson BJ, et al**.** Endoscopically directed middle meatal cultures versus maxillary sinus taps in acute bacterial maxillary rhinosinusitis: a meta-analysis. Otolaryngol Head Neck Surg. 2006;134(1):3-9.
50. Berger MY, van der Velden JJ, Lijmer JG, et al**.** Abdominal symptoms: do they predict gallstones? A systematic review. Scand J Gastroenterol. 2000;35(1):70-6.
51. Berman DS, Kiat H, Van Train KF, et al**.** Comparison of SPECT using technetium-99m agents and thallium-201 and PET for the assessment of myocardial perfusion and viability. Am J Cardiol. 1990;66(13):72E-9E.
52. Berner MM, Kriston L, Bentele M, et al**.** The alcohol use disorders identification test for detecting at-risk drinking: a systematic review and meta-analysis. J Stud Alcohol Drugs. 2007;68(3):461-73.
53. Bhardwaj A, Hollenbeak CS, Pooran N, et al**.** A meta-analysis of the diagnostic accuracy of esophageal capsule endoscopy for Barrett's esophagus in patients with gastroesophageal reflux disease. Am J Gastroenterol. 2009;104(6):1533-9.
54. Bipat S, Glas AS, Slors FJ, et al**.** Rectal cancer: local staging and assessment of lymph node involvement with endoluminal US, CT, and MR imaging—a meta-analysis. Radiology. 2004;232(3):773-83.

55. Bipat S, Glas AS, van der Velden J, et al**.** Computed tomography and magnetic resonance imaging in staging of uterine cervical carcinoma: a systematic review. Gynecol Oncol. 2003;91(1):59-66.
56. Bipat S, Phoa SS, van Delden OM, et al. Ultrasonography, computed tomography and magnetic resonance imaging for diagnosis and determining resectability of pancreatic adenocarcinoma: a meta-analysis. J Comput Assist Tomogr. 2005;29(4):438-45.
57. Bipat S, van Leeuwen MS, Comans EF, et al. Colorectal liver metastases: CT, MR imaging, and PET for diagnosis—meta-analysis. Radiology. 2005;237(1):123-31.
58. Birim O, Kappetein AP, Stijnen T, et al**.** Meta-analysis of positron emission tomographic and computed tomographic imaging in detecting mediastinal lymph node metastases in nonsmall cell lung cancer. Ann Thorac Surg. 2005;79(1):375-82.
59. Blacksell SD, Doust JA, Newton PN, et al**.** A systematic review and meta-analysis of the diagnostic accuracy of rapid immunochromatographic assays for the detection of dengue virus IgM antibodies during acute infection. Trans R Soc Trop Med Hyg. 2006;100(8):775-84.
60. Blakeley DD, Oddone EZ, Hasselblad V, et al**.** Noninvasive carotid artery testing. A meta-analytic review. Ann Intern Med. 1995;122(5):360-7.
61. Blaufox MD, Middleton ML, Bongiovanni J, et al**.** Cost efficacy of the diagnosis and therapy of renovascular hypertension. J Nucl Med. 1996;37(1):171-7.
62. Bonis PA, Ioannidis JP, Cappelleri JC, et al**.** Correlation of biochemical response to interferon alfa with histological improvement in hepatitis C: a meta-analysis of diagnostic test characteristics. Hepatology. 1997;26(4):1035-44.
63. Booth CM, Boone RH, Tomlinson G, et al**.** Is this patient dead, vegetative, or severely neurologically impaired? Assessing outcome for comatose survivors of cardiac arrest. JAMA. 2004;291(7):870-9.
64. Brealey S, Scally A, Hahn S, et al**.** Accuracy of radiographer plain radiograph reporting in clinical practice: a meta-analysis. Clin Radiol. 2005;60(2):232-41.
65. Brealey S, Scally A, Hahn S, et al**.** Accuracy of radiographers red dot or triage of accident and emergency radiographs in clinical practice: a systematic review. Clin Radiol. 2006;61(7):604-15.
66. Brennan ME, Houssami N, Lord S, et al. Magnetic resonance imaging screening of the contralateral breast in women with newly diagnosed breast cancer: systematic review and meta-analysis of incremental cancer detection and impact on surgical management. J Clin Oncol. 2009;27(33):5640-9.
67. Brietzke SE, Katz ES, Roberson DW**.** Can history and physical examination reliably diagnose pediatric obstructive sleep apnea/hypopnea syndrome? A systematic review of the literature. Otolaryngol Head Neck Surg. 2004;131(6):827-32.
68. Broekmans FJ, Kwee J, Hendriks DJ, et al**.** A systematic review of tests predicting ovarian reserve and IVF outcome. Hum Reprod Update. 2006;12(6):685-718.
69. Broer SL, Mol BW, Hendriks D, et al**.** The role of antimullerian hormone in prediction of outcome after IVF: comparison with the antral follicle count. Fertil Steril. 2009;91(3):705-14.
70. Brouwer J, Hooft L, Hoekstra OS, et al. Systematic review: accuracy of imaging tests in the diagnosis of recurrent laryngeal carcinoma after radiotherapy. Head Neck. 2008;30(7):889-97.
71. Brown DL, Doubilet PM**.** Transvaginal sonography for diagnosing ectopic pregnancy: positivity criteria and performance characteristics. J Ultrasound Med. 1994;13(4):259-66.
72. Brown MD, Lau J, Nelson RD, et al**.** Turbidimetric D-dimer test in the diagnosis of pulmonary embolism: a metaanalysis. Clin Chem. 2003;49(11):1846-53.

73. Brown MD, Rowe BH, Reeves MJ, et al. The accuracy of the enzyme-linked immunosorbent assay D-dimer test in the diagnosis of pulmonary embolism: a meta-analysis. Ann Emerg Med. 2002;40(2):133-44.
74. Bruyninckx R, Aertgeerts B, Bruyninckx P, et al. Signs and symptoms in diagnosing acute myocardial infarction and acute coronary syndrome: a diagnostic meta-analysis. Br J Gen Pract. 2008;58(547):105-11.
75. Buntinx F, Wauters H. The diagnostic value of macroscopic haematuria in diagnosing urological cancers: a meta-analysis. Fam Pract. 1997;14(1):63-8.
76. Burnside PR, Brown MD, Kline JA. Systematic review of emergency physician-performed ultrasonography for lower-extremity deep vein thrombosis. Acad Emerg Med. 2008;15(6):493-8.
77. Bwanga F, Hoffner S, Haile M, et al. Direct susceptibility testing for multi drug resistant tuberculosis: a meta-analysis. BMC Infect Dis. 2009;9:67.
78. Cabana MD, Alavi A, Berlin JA, et al. Morphine-augmented hepatobiliary scintigraphy: a meta-analysis. Nucl Med Commun. 1995;16(12):1068-71.
79. Campens D, Buntinx F. Selecting the best renal function tests. A meta-analysis of diagnostic studies. Int J Technol Assess Health Care. 1997;13(2):343-56.
80. Carlson KJ, Skates SJ, Singer DE. Screening for ovarian cancer. Ann Intern Med. 1994;121(2):124-32.
81. Carroll T, Raff H, Findling JW. Late-night salivary cortisol for the diagnosis of Cushing syndrome: a meta-analysis. Endocr Pract. 2009;15(4):335-42.
82. Carter BG, Butt W. Review of the use of somatosensory evoked potentials in the prediction of outcome after severe brain injury. Crit Care Med. 2001;29(1):178-86.
83. Carter BG, Butt W. Are somatosensory evoked potentials the best predictor of outcome after severe brain injury? A systematic review. Intensive Care Med. 2005;31(6):765-75.
84. Cascini GL, De Palma D, Matteucci F, et al. Fever of unknown origin, infection of subcutaneous devices, brain abscesses and endocarditis. Nucl Med Commun. 2006;27(3):213-22.
85. Castilla-Rilo J, Lopez-Arrieta J, Bermejo-Pareja F, et al. Instrumental activities of daily living in the screening of dementia in population studies: a systematic review and meta-analysis. Int J Geriatr Psychiatry. 2007;22(9):829-36.
86. Cavallazzi R, Nair A, Vasu T, et al. Natriuretic peptides in acute pulmonary embolism: a systematic review. Intensive Care Med. 2008;34(12):2147-56.
87. Cepoiu M, McCusker J, Cole MG, et al. Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. J Gen Intern Med. 2008;23(1):25-36.
88. Chalco JP, Huicho L, Alamo C, et al. Accuracy of clinical pallor in the diagnosis of anaemia in children: a meta-analysis. BMC Pediatr. 2005;5:46.
89. Chan BK, Melnikow J, Slee CA, et al. Posttreatment human papillomavirus testing for recurrent cervical intraepithelial neoplasia: a systematic review. Am J Obstet Gynecol. 2009;200(4):422 e1-9.
90. Chaparro M, Gisbert JP, Del Campo L, et al. Accuracy of computed tomographic colonography for the detection of polyps and colorectal tumors: a systematic review and meta-analysis. Digestion. 2009;80(1):1-17.
91. Chappell ET, Moure FC, Good MC. Comparison of computed tomographic angiography with digital subtraction angiography in the diagnosis of cerebral aneurysms: a meta-analysis. Neurosurgery. 2003;52(3):624-31; discussion 30-1.
92. Chappuis F, Rijal S, Soto A, et al. A meta-analysis of the diagnostic performance of the direct agglutination test and rK39 dipstick for visceral leishmaniasis. BMJ. 2006;333(7571):723.

93. Chen SC, Bravata DM, Weil E, et al**.** A comparison of dermatologists' and primary care physicians' accuracy in diagnosing melanoma: a systematic review. Arch Dermatol. 2001;137(12):1627-34.
94. Cher DJ, Conwell JA, Mandel JS**.** MRI for detecting silicone breast implant rupture: meta-analysis and implications. Ann Plast Surg. 2001;47(4):367-80.
95. Chien PF, Arnott N, Gordon A, et al**.** How useful is uterine artery Doppler flow velocimetry in the prediction of pre-eclampsia, intrauterine growth retardation and perinatal death? An overview. BJOG. 2000;107(2):196-208.
96. Chien PF, Khan KS, Ogston S, et al**.** The diagnostic accuracy of cervico-vaginal fetal fibronectin in predicting preterm delivery: an overview. Br J Obstet Gynaecol. 1997;104(4):436-44.
97. Chin AS, Goldman LE, Eisenberg MJ**.** Functional testing after coronary artery bypass graft surgery: a meta-analysis. Can J Cardiol. 2003;19(7):802-8.
98. Choi HK, Liu S, Merkel PA, et al**.** Diagnostic performance of antineutrophil cytoplasmic antibody tests for idiopathic vasculitides: metaanalysis with a focus on antimyeloperoxidase antibodies. J Rheumatol. 2001;28(7):1584-90.
99. Christou MA, Siontis GC, Katritsis DG, et al**.** Meta-analysis of fractional flow reserve versus quantitative coronary angiography and noninvasive imaging for evaluation of myocardial ischemia. Am J Cardiol. 2007;99(4):450-6.
100. Chua AE, Ridley LJ**.** Diagnostic accuracy of CT angiography in acute gastrointestinal bleeding. J Med Imaging Radiat Oncol. 2008;52(4):333-8.
101. Chun AA, McGee SR**.** Bedside diagnosis of coronary artery disease: a systematic review. Am J Med. 2004;117(5):334-43.
102. Cinnella G, Dambrosio M, Brienza N, et a;**.** Transesophageal echocardiography for diagnosis of traumatic aortic injury: an appraisal of the evidence. J Trauma. 2004;57(6):1246-55.
103. Clark TJ, Mann CH, Shah N, et al**.** Accuracy of outpatient endometrial biopsy in the diagnosis of endometrial hyperplasia. Acta Obstet Gynecol Scand. 2001;80(9):784-93.
104. Clark TJ, ter Riet G, Coomarasamy A, et al**.** Bias associated with delayed verification in test accuracy studies: accuracy of tests for endometrial hyperplasia may be much higher than we think! BMC Med. 2004;2:18.
105. Clark TJ, Voit D, Gupta JK, et al**.** Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review. JAMA. 2002;288(13):1610-21.
106. Clarke CE, Davies P**.** Systematic review of acute levodopa and apomorphine challenge tests in the diagnosis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry. 2000;69(5):590-4.
107. Clerico A, Fontana M, Zyw L, et al**.** Comparison of the diagnostic accuracy of brain natriuretic peptide (BNP) and the N-terminal part of the propeptide of BNP immunoassays in chronic and acute heart failure: a systematic review. Clin Chem. 2007;53(5):813-22.
108. Cnossen JS, Morris RK, ter Riet G, et al. Use of uterine artery Doppler ultrasonography to predict pre-eclampsia and intrauterine growth restriction: a systematic review and bivariable meta-analysis. CMAJ. 2008;178(6):701-11.
109. Cnossen JS, Vollebregt KC, de Vrieze N, et al. Accuracy of mean arterial pressure and blood pressure measurements in predicting pre-eclampsia: systematic review and meta-analysis. BMJ. 2008;336(7653):1117-20.
110. Colin C, Lanoir D, Touzet S, et al**.** Sensitivity and specificity of third-generation hepatitis C virus antibody detection assays: an analysis of the literature. J Viral Hepat. 2001;8(2):87-95.

111. Colli A, Fraquelli M, Casazza G, et al. Accuracy of ultrasonography, spiral CT, magnetic resonance, and alpha-fetoprotein in diagnosing hepatocellular carcinoma: a systematic review. Am J Gastroenterol. 2006;101(3):513-23.
112. Collins JA, Barnhart KT, Schlegel PN**.** Do sperm DNA integrity tests predict pregnancy with in vitro fertilization? Fertil Steril. 2008;89(4):823-31.
113. Conde-Agudelo A, Kafury-Goeta AC**.** Triple-marker test as screening for Down syndrome: a meta-analysis. Obstet Gynecol Surv. 1998;53(6):369-76.
114. Conde-Agudelo A, Villar J, Lindheimer M**.** World Health Organization systematic review of screening tests for preeclampsia. Obstet Gynecol. 2004;104(6):1367-91.
115. Contopoulos-Ioannidis DG, Ioannidis JP**.** Maternal cell-free viremia in the natural history of perinatal HIV-1 transmission: a meta-analysis. J Acquir Immune Defic Syndr Hum Retrovirol. 1998;18(2):126-35.
116. Cook DJ, Fitzgerald JM, Guyatt GH, et al**.** Evaluation of the protected brush catheter and bronchoalveolar lavage in the diagnosis of nosocomial pneumonia. J Intensive Care Med. 1991;6(4):196-205.
117. Cook RL, Hutchison SL, Ostergaard L, et al**.** Systematic review: noninvasive testing for Chlamydia trachomatis and Neisseria gonorrhoeae. Ann Intern Med. 2005;142(11):914-25.
118. Cooke G, Doust J, Sanders S**.** Is pulse palpation helpful in detecting atrial fibrillation? A systematic review. J Fam Pract. 2006;55(2):130-4.
119. Cote AM, Brown MA, Lam E, et al. Diagnostic accuracy of urinary spot protein:creatinine ratio for proteinuria in hypertensive pregnant women: systematic review. BMJ. 2008;336(7651):1003-6.
120. Coutance G, Le Page O, Lo T, et al. Prognostic value of brain natriuretic peptide in acute pulmonary embolism. Crit Care. 2008;12(4):R109.
121. Crane JM**.** Factors predicting labor induction success: a critical analysis. Clin Obstet Gynecol. 2006;49(3):573-84.
122. Crane JM, Hutchens D**.** Transvaginal sonographic measurement of cervical length to predict preterm birth in asymptomatic women at increased risk: a systematic review. Ultrasound Obstet Gynecol. 2008;31(5):579-87.
123. Crawford R, Walley G, Bridgman S, et al**.** Magnetic resonance imaging versus arthroscopy in the diagnosis of knee pathology, concentrating on meniscal lesions and ACL tears: a systematic review. Br Med Bull. 2007;84:5-23.
124. Cremonini F, Wise J, Moayyedi P, et al**.** Diagnostic and therapeutic use of proton pump inhibitors in non-cardiac chest pain: a metaanalysis. Am J Gastroenterol. 2005;100(6):1226-32.
125. Critchley J, Bates I**.** Haemoglobin colour scale for anaemia diagnosis where there is no laboratory: a systematic review. Int J Epidemiol. 2005;34(6):1425-34.
126. Crobach MJ, Dekkers OM, Wilcox MH, et al**.** European Society of Clinical Microbiology and Infectious Diseases (ESCMID): data review and recommendations for diagnosing Clostridium difficile-infection (CDI). Clin Microbiol Infect. 2009;15(12):1053-66.
127. Cronin P, Dwamena BA, Kelly AM, et al**.** Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. Eur Radiol. 2008;18(9):1840-56.
128. Cronin P, Dwamena BA, Kelly AM, et al**.** Solitary pulmonary nodules: meta-analytic comparison of cross-sectional imaging modalities for diagnosis of malignancy. Radiology. 2008;246(3):772-82.
129. Cruciani M, Marcati P, Malena M, et al**.** Meta-analysis of diagnostic procedures for Pneumocystis carinii pneumonia in HIV-1-infected patients. Eur Respir J. 2002;20(4):982-9.

130. Cruciani M, Nardi S, Malena M, et al. Systematic review of the accuracy of the ParaSight-F test in the diagnosis of Plasmodium falciparum malaria. Med Sci Monit. 2004;10(7):MT81-8.
131. Cruciani M, Scarparo C, Malena M, et al. Meta-analysis of BACTEC MGIT 960 and BACTEC 460 TB, with or without solid media, for detection of mycobacteria. J Clin Microbiol. 2004;42(5):2321-5.
132. Cuckle H, van Oudgaarden ED, Mason G, et al. Taking account of vaginal bleeding in screening for Down's syndrome. Br J Obstet Gynaecol. 1994;101(11):948-53.
133. Cueto SM, Cavanaugh SH, Benenson RS, et al. Computed tomography scan versus ventilation-perfusion lung scan in the detection of pulmonary embolism. J Emerg Med. 2001;21(2):155-64.
134. Curvers WL, van den Broek FJ, Reitsma JB, et al. Systematic review of narrow-band imaging for the detection and differentiation of abnormalities in the esophagus and stomach (with video). Gastrointest Endosc. 2009;69(2):307-17.
135. Cury RC, Feutchner G, Pena CS, et al. Acute chest pain imaging in the emergency department with cardiac computed tomography angiography. J Nucl Cardiol. 2008;15(4):564-75.
136. Cuzick J, Arbyn M, Sankaranarayanan R, et al. Overview of human papillomavirus-based and other novel options for cervical cancer screening in developed and developing countries. Vaccine. 2008;26 Suppl 10:K29-41.
137. Cuzick J, Clavel C, Petry KU, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. Int J Cancer. 2006;119(5):1095-101.
138. Dales RE, Stark RM, Raman S. Computed tomography to stage lung cancer. Approaching a controversy using meta-analysis. Am Rev Respir Dis. 1990;141(5 Pt 1):1096-101.
139. Daley P, Thomas S, Pai M. Nucleic acid amplification tests for the diagnosis of tuberculous lymphadenitis: a systematic review. Int J Tuberc Lung Dis. 2007;11(11):1166-76.
140. Danias PG, Roussakis A, Ioannidis JP. Diagnostic performance of coronary magnetic resonance angiography as compared against conventional X-ray angiography: a meta-analysis. J Am Coll Cardiol. 2004;44(9):1867-76.
141. D'Arcy CA, McGee S. The rational clinical examination. Does this patient have carpal tunnel syndrome? JAMA. 2000;283(23):3110-7.
142. Davison SP, Clifton MS, Kauffman L, et al. Sentinel node biopsy for the detection of head and neck melanoma: a review. Ann Plast Surg. 2001;47(2):206-11.
143. de Albuquerque Fonseca L, Picano E. Comparison of dipyridamole and exercise stress echocardiography for detection of coronary artery disease (a meta-analysis). Am J Cardiol. 2001;87(10):1193-6; A4.
144. De Bernardinis M, Violi V, Roncoroni L, et al. Discriminant power and information content of Ranson's prognostic signs in acute pancreatitis: a meta-analytic study. Crit Care Med. 1999;27(10):2272-83.
145. de Bondt RB, Nelemans PJ, Hofman PA, et al. Detection of lymph node metastases in head and neck cancer: a meta-analysis comparing US, USgFNAC, CT and MR imaging. Eur J Radiol. 2007;64(2):266-72.
146. de Bruyn G, Graviss EA. A systematic review of the diagnostic accuracy of physical examination for the detection of cirrhosis. BMC Med Inform Decis Mak. 2001;1:6.
147. de Jaeger A, Litalien C, Lacroix J, et al. Protected specimen brush or bronchoalveolar lavage to diagnose bacterial nosocomial pneumonia in ventilated adults: a meta-analysis. Crit Care Med. 1999;27(11):2548-60.

148. de Jesus JO, Parker L, Frangos AJ, et al. Accuracy of MRI, MR arthrography, and ultrasound in the diagnosis of rotator cuff tears: a meta-analysis. AJR Am J Roentgenol. 2009;192(6):1701-7.
149. de Kroon CD, de Bock GH, Dieben SW, et al. Saline contrast hysterosonography in abnormal uterine bleeding: a systematic review and meta-analysis. BJOG. 2003;110(10):938-47.
150. de Lorijn F, Kremer LC, Reitsma JB, et al. Diagnostic tests in Hirschsprung disease: a systematic review. J Pediatr Gastroenterol Nutr. 2006;42(5):496-505.
151. de Sousa MR, Morillo CA, Rabelo FT, et al. Non-sustained ventricular tachycardia as a predictor of sudden cardiac death in patients with left ventricular dysfunction: a meta-analysis. Eur J Heart Fail. 2008;10(10):1007-14.
152. de Vries SO, Hunink MG, 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(4):361-9.
153. Debrey SM, Yu H, Lynch JK, et al. Diagnostic accuracy of magnetic resonance angiography for internal carotid artery disease: a systematic review and meta-analysis. Stroke. 2008;39(8):2237-48.
154. Delgado-Bolton RC, Fernandez-Perez C, Gonzalez-Mate A, et al. Meta-analysis of the performance of 18F-FDG PET in primary tumor detection in unknown primary tumors. J Nucl Med. 2003;44(8):1301-14.
155. Denham DW, Norman J. Cost-effectiveness of preoperative sestamibi scan for primary hyperparathyroidism is dependent solely upon the surgeon's choice of operative procedure. J Am Coll Surg. 1998;186(3):293-305.
156. Des Guetz G, Uzzan B, Nicolas P, et al. Is sentinel lymph node mapping in colorectal cancer a future prognostic factor? A meta-analysis. World J Surg. 2007;31(6):1304-12.
157. Detrano R, Gianrossi R, Mulvihill D, et al. Exercise-induced ST segment depression in the diagnosis of multivessel coronary disease: a meta analysis. J Am Coll Cardiol. 1989;14(6):1501-8.
158. Detrano R, Janosi A, Lyons KP, et al. Factors affecting sensitivity and specificity of a diagnostic test: the exercise thallium scintigram. Am J Med. 1988;84(4):699-710.
159. Detsky ME, McDonald DR, Baerlocher MO, et al. Does this patient with headache have a migraine or need neuroimaging? JAMA. 2006;296(10):1274-83.
160. Deville WL, van der Windt DA, Dzaferagic A, et al. The test of Lasegue: systematic review of the accuracy in diagnosing herniated discs. Spine (Phila Pa 1976). 2000;25(9):1140-7.
161. Deville WL, Yzermans JC, van Duijn NP, et al. The urine dipstick test useful to rule out infections. A meta-analysis of the accuracy. BMC Urol. 2004;4:4.
162. Devous MD, Sr., Thisted RA, Morgan GF, et al. SPECT brain imaging in epilepsy: a meta-analysis. J Nucl Med. 1998;39(2):285-93.
163. Di Fabio RP. Meta-analysis of the sensitivity and specificity of platform posturography. Arch Otolaryngol Head Neck Surg. 1996;122(2):150-6.
164. Di Nisio M, Squizzato A, Rutjes AW, et al. Diagnostic accuracy of D-dimer test for exclusion of venous thromboembolism: a systematic review. J Thromb Haemost. 2007;5(2):296-304.
165. Di Tanna GL, Berti E, Stivanello E, et al. Informative value of clinical research on multislice computed tomography in the diagnosis of coronary artery disease: A systematic review. Int J Cardiol. 2008;130(3):386-404.
166. Dijkhuizen FP, Mol BW, Brolmann HA, et al. The accuracy of endometrial sampling in the diagnosis of patients with endometrial carcinoma and hyperplasia: a meta-analysis. Cancer. 2000;89(8):1765-72.

167. Dinh MT, Abad CL, Safdar N. Diagnostic accuracy of the physical examination and imaging tests for osteomyelitis underlying diabetic foot ulcers: meta-analysis. Clin Infect Dis. 2008;47(4):519-27.
168. Dinis-Ribeiro M, Yamaki G, Miki K, et al. Meta-analysis on the validity of pepsinogen test for gastric carcinoma, dysplasia or chronic atrophic gastritis screening. J Med Screen. 2004;11(3):141-7.
169. Dinnes J, Deeks J, Kunst H, et al. A systematic review of rapid diagnostic tests for the detection of tuberculosis infection. Health Technol Assess. 2007;11(3):1-196.
170. Dodd SR, Lancaster GA, Craig JV, et al. In a systematic review, infrared ear thermometry for fever diagnosis in children finds poor sensitivity. J Clin Epidemiol. 2006;59(4):354-7.
171. Dong MJ, Liu ZF, Zhao K, et al. Value of 18F-FDG-PET/PET-CT in differentiated thyroid carcinoma with radioiodine-negative whole-body scan: a meta-analysis. Nucl Med Commun. 2009;30(8):639-50.
172. Dong MJ, Zhao K, Lin XT, et al. Role of fluorodeoxyglucose-PET versus fluorodeoxyglucose-PET/computed tomography in detection of unknown primary tumor: a meta-analysis of the literature. Nucl Med Commun. 2008;29(9):791-802.
173. Doobay AV, Anand SS. Sensitivity and specificity of the ankle-brachial index to predict future cardiovascular outcomes: a systematic review. Arterioscler Thromb Vasc Biol. 2005;25(7):1463-9.
174. Doria AS, Moineddin R, Kellenberger CJ, et al. US or CT for Diagnosis of Appendicitis in Children and Adults? A Meta-Analysis. Radiology. 2006;241(1):83-94.
175. Dougall NJ, Bruggink S, Ebmeier KP. Systematic review of the diagnostic accuracy of 99mTc-HMPAO-SPECT in dementia. Am J Geriatr Psychiatry. 2004;12(6):554-70.
176. Doust JA, Glasziou PP, Pietrzak E, et al. A systematic review of the diagnostic accuracy of natriuretic peptides for heart failure. Arch Intern Med. 2004;164(18):1978-84.
177. Dove SB. Radiographic diagnosis of dental caries. J Dent Educ. 2001;65(10):985-90.
178. Dowling S, Spooner CH, Liang Y, et al. Accuracy of Ottawa Ankle Rules to exclude fractures of the ankle and midfoot in children: a meta-analysis. Acad Emerg Med. 2009;16(4):277-87.
179. Downer MC, Moles DR, Palmer S, et al. A systematic review of test performance in screening for oral cancer and precancer. Oral Oncol. 2004;40(3):264-73.
180. Dubin MG, Ebert CS, Coffey CS, et al. Concordance of middle meatal swab and maxillary sinus aspirate in acute and chronic sinusitis: a meta-analysis. Am J Rhinol. 2005;19(5):462-70.
181. Dumler JS. Molecular diagnosis of Lyme disease: review and meta-analysis. Mol Diagn. 2001;6(1):1-11.
182. Dwamena BA, Sonnad SS, Angobaldo JO, et al. Metastases from non-small cell lung cancer: mediastinal staging in the 1990s—meta-analytic comparison of PET and CT. Radiology. 1999;213(2):530-6.
183. East JE, Tan EK, Bergman JJ, et al. Meta-analysis: narrow band imaging for lesion characterization in the colon, oesophagus, duodenal ampulla and lung. Aliment Pharmacol Ther. 2008;28(7):854-67.
184. Ebell MH, Flewelling D, Flynn CA. A systematic review of troponin T and I for diagnosing acute myocardial infarction. J Fam Pract. 2000;49(6):550-6.
185. Ebell MH, White LL, Casault T. A systematic review of the history and physical examination to diagnose influenza. J Am Board Fam Pract. 2004;17(1):1-5.
186. Ebell MH, White LL, Weismantel D. A systematic review of troponin T and I values as a prognostic tool for patients with chest pain. J Fam Pract. 2000;49(8):746-53.
187. Eiberg JP, Lundorf E, Thomsen C, et al. Peripheral vascular surgery and magnetic resonance arteriography—a review. Eur J Vasc Endovasc Surg. 2001;22(5):396-402.

188. Elamin MB, Murad MH, Mullan R, et al. Accuracy of diagnostic tests for Cushing's syndrome: a systematic review and metaanalyses. J Clin Endocrinol Metab. 2008;93(5):1553-62.
189. Engelbrecht MR, Jager GJ, Laheij RJ, et al**.** Local staging of prostate cancer using magnetic resonance imaging: a meta-analysis. Eur Radiol. 2002;12(9):2294-302.
190. Engels EA, Terrin N, Barza M, et al**.** Meta-analysis of diagnostic tests for acute sinusitis. J Clin Epidemiol. 2000;53(8):852-62.
191. Engman ML**.** An update on EBCT (Ultrafast CT) scans for coronary artery disease. J Insur Med. 1998;30(3):175-9.
192. Eustatia-Rutten CF, Smit JW, Romijn JA, et al. Diagnostic value of serum thyroglobulin measurements in the follow-up of differentiated thyroid carcinoma, a structured meta-analysis. Clin Endocrinol (Oxf). 2004;61(1):61-74.
193. Evennett N, Alexander N, Petrov M, et al**.** A systematic review of serologic tests in the diagnosis of necrotizing enterocolitis. J Pediatr Surg. 2009;44(11):2192-201.
194. Evennett NJ, Petrov MS, Mittal A, et al**.** Systematic review and pooled estimates for the diagnostic accuracy of serological markers for intestinal ischemia. World J Surg. 2009;33(7):1374-83.
195. Ewald B, Attia J**.** Which test to detect microalbuminuria in diabetic patients? A systematic review. Aust Fam Physician. 2004;33(7):565-7, 71.
196. Ewald B, Ewald D, Thakkinstian A, et al**.** Meta-analysis of B type natriuretic peptide and N-terminal pro B natriuretic peptide in the diagnosis of clinical heart failure and population screening for left ventricular systolic dysfunction. Intern Med J. 2008;38(2):101-13.
197. Fahey MT, Irwig L, Macaskill P**.** Meta-analysis of Pap test accuracy. Am J Epidemiol. 1995;141(7):680-9.
198. Falagas ME, Kazantzi MS, Bliziotis IA**.** Comparison of utility of blood cultures from intravascular catheters and peripheral veins: a systematic review and decision analysis. J Med Microbiol. 2008;57(Pt 1):1-8.
199. Falk G, Fahey T**.** C-reactive protein and community-acquired pneumonia in ambulatory care: systematic review of diagnostic accuracy studies. Fam Pract. 2009;26(1):10-21.
200. Fancher TL, White RH, Kravitz RL**.** Combined use of rapid D-dimer testing and estimation of clinical probability in the diagnosis of deep vein thrombosis: systematic review. BMJ. 2004;329(7470):821.
201. Faron G, Boulvain M, Irion O, et al**.** Prediction of preterm delivery by fetal fibronectin: a meta-analysis. Obstet Gynecol. 1998;92(1):153-8.
202. Farquhar C, Ekeroma A, Furness S, et al**.** A systematic review of transvaginal ultrasonography, sonohysterography and hysteroscopy for the investigation of abnormal uterine bleeding in premenopausal women. Acta Obstet Gynecol Scand. 2003;82(6):493-504.
203. Farrell T, Chien PF, Gordon A**.** Intrapartum umbilical artery Doppler velocimetry as a predictor of adverse perinatal outcome: a systematic review. Br J Obstet Gynaecol. 1999;106(8):783-92.
204. Feng Y, Schlosser FJ, Sumpio BE**.** The Semmes Weinstein monofilament examination as a screening tool for diabetic peripheral neuropathy. J Vasc Surg. 2009;50(3):675-82, 82 e1.
205. Fernandez Alvarez JR, Amess PN, Gandhi RS, et al**.** Diagnostic value of subependymal pseudocysts and choroid plexus cysts on neonatal cerebral ultrasound: a meta-analysis. Arch Dis Child Fetal Neonatal Ed. 2009;94(6):F443-6.
206. Fischer BM, Mortensen J, Hojgaard L**.** Positron emission tomography in the diagnosis and staging of lung cancer: a systematic, quantitative review. Lancet Oncol. 2001;2(11):659-66.

207. Fleischmann KE, Hunink MG, Kuntz KM, et al**.** Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA. 1998;280(10):913-20.
208. Flood RG, Badik J, Aronoff SC**.** The utility of serum C-reactive protein in differentiating bacterial from nonbacterial pneumonia in children: a meta-analysis of 1230 children. Pediatr Infect Dis J. 2008;27(2):95-9.
209. Flores LL, Pai M, Colford JM, Jr., et al**.** In-house nucleic acid amplification tests for the detection of Mycobacterium tuberculosis in sputum specimens: meta-analysis and meta-regression. BMC Microbiol. 2005;5:55.
210. Ford AC, Chey WD, Talley NJ, et al**.** Yield of diagnostic tests for celiac disease in individuals with symptoms suggestive of irritable bowel syndrome: systematic review and meta-analysis. Arch Intern Med. 2009;169(7):651-8.
211. Ford AC, Ching E, Moayyedi P**.** Meta-analysis: yield of diagnostic tests for coeliac disease in dyspepsia. Aliment Pharmacol Ther. 2009;30(1):28-36.
212. Ford AC, Spiegel BM, Talley NJ, et al**.** Small intestinal bacterial overgrowth in irritable bowel syndrome: systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2009;7(12):1279-86.
213. Ford AC, Veldhuyzen van Zanten SJ, Rodgers CC, et al**.** Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. Gut. 2008;57(11):1545-53.
214. Fraile M, Rull M, Julian FJ, et al. Sentinel node biopsy as a practical alternative to axillary lymph node dissection in breast cancer patients: an approach to its validity. Ann Oncol. 2000;11(6):701-5.
215. Fransen GA, Janssen MJ, Muris JW, et al**.** Meta-analysis: the diagnostic value of alarm symptoms for upper gastrointestinal malignancy. Aliment Pharmacol Ther. 2004;20(10):1045-52.
216. Fraquelli M, Colli A, Casazza G, et al. Role of US in detection of Crohn disease: meta-analysis. Radiology. 2005;236(1):95-101.
217. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of transient elastography for the staging of liver fibrosis: a meta-analysis. Gastroenterology. 2008;134(4):960-74.
218. Gao H, McDonnell A, Harrison DA, et al. Systematic review and evaluation of physiological track and trigger warning systems for identifying at-risk patients on the ward. Intensive Care Med. 2007;33(4):667-79.
219. Garrow D, Miller S, Sinha D, et al. Endoscopic ultrasound: a meta-analysis of test performance in suspected biliary obstruction. Clin Gastroenterol Hepatol. 2007;5(5):616-23.
220. Garzon PP, Eisenberg MJ**.** Functional testing for the detection of restenosis after percutaneous transluminal coronary angioplasty: a meta-analysis. Can J Cardiol. 2001;17(1):41-8.
221. Geersing GJ, Janssen KJ, Oudega R, et al. Excluding venous thromboembolism using point of care D-dimer tests in outpatients: a diagnostic meta-analysis. BMJ. 2009;339:b2990.
222. Gehi AK, Stein RH, Metz LD, et al**.** Microvolt T-wave alternans for the risk stratification of ventricular tachyarrhythmic events: a meta-analysis. J Am Coll Cardiol. 2005;46(1):75-82.
223. Geisser ME, Ranavaya M, Haig AJ, et al. A meta-analytic review of surface electromyography among persons with low back pain and normal, healthy controls. J Pain. 2005;6(11):711-26.
224. Geleijnse ML, Krenning BJ, Soliman OI, et al**.** Dobutamine stress echocardiography for the detection of coronary artery disease in women. Am J Cardiol. 2007;99(5):714-7.

225. Geleijnse ML, Krenning BJ, van Dalen BM, et al. Factors affecting sensitivity and specificity of diagnostic testing: dobutamine stress echocardiography. J Am Soc Echocardiogr. 2009;22(11):1199-208.
226. Geomini P, Bremer G, Kruitwagen R, et al**.** Diagnostic accuracy of frozen section diagnosis of the adnexal mass: a metaanalysis. Gynecol Oncol. 2005;96(1):1-9.
227. Gerdes LU, Jorgensen PE, Nexo E, et al**.** C-reactive protein and bacterial meningitis: a meta-analysis. Scand J Clin Lab Invest. 1998;58(5):383-93.
228. Gianrossi R, Detrano R, Colombo A, et al**.** Cardiac fluoroscopy for the diagnosis of coronary artery disease: a meta analytic review. Am Heart J. 1990;120(5):1179-88.
229. Gibreel A, Maheshwari A, Bhattacharya S, et al**.** Ultrasound tests of ovarian reserve; a systematic review of accuracy in predicting fertility outcomes. Hum Fertil (Camb). 2009;12(2):95-106.
230. Gilbody S, Richards D, Brealey S, et al**.** Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med. 2007;22(11):1596-602.
231. Gill CJ, Lau J, Gorbach SL, et al**.** Diagnostic accuracy of stool assays for inflammatory bacterial gastroenteritis in developed and resource-poor countries. Clin Infect Dis. 2003;37(3):365-75.
232. Gisbert JP, Abraira V**.** Accuracy of Helicobacter pylori diagnostic tests in patients with bleeding peptic ulcer: a systematic review and meta-analysis. Am J Gastroenterol. 2006;101(4):848-63.
233. Gisbert JP, de la Morena F, Abraira V**.** Accuracy of monoclonal stool antigen test for the diagnosis of H. pylori infection: a systematic review and meta-analysis. Am J Gastroenterol. 2006;101(8):1921-30.
234. Gisbert JP, Pajares JM**.** Diagnosis of Helicobacter pylori infection by stool antigen determination: a systematic review. Am J Gastroenterol. 2001;96(10):2829-38.
235. Gisbert JP, Pajares JM**.** Stool antigen test for the diagnosis of Helicobacter pylori infection: a systematic review. Helicobacter. 2004;9(4):347-68.
236. Glas AS, Roos D, Deutekom M, et al**.** Tumor markers in the diagnosis of primary bladder cancer. A systematic review. J Urol. 2003;169(6):1975-82.
237. Goldstein LB, Simel DL**.** Is this patient having a stroke? JAMA. 2005;293(19):2391-402.
238. Goodacre S, Sampson FC, Sutton AJ, et al**.** Variation in the diagnostic performance of D-dimer for suspected deep vein thrombosis. QJM. 2005;98(7):513-27.
239. Goodacre S, Sutton AJ, Sampson FC**.** Meta-analysis: The value of clinical assessment in the diagnosis of deep venous thrombosis. Ann Intern Med. 2005;143(2):129-39.
240. Goodman CM, Cohen V, Thornby J, et al**.** The life span of silicone gel breast implants and a comparison of mammography, ultrasonography, and magnetic resonance imaging in detecting implant rupture: a meta-analysis. Ann Plast Surg. 1998;41(6):577-85; discussion 85-6.
241. Goodman CS, Hur JY, Adajar MA, et al**.** How well does CT predict the need for laparotomy in hemodynamically stable patients with penetrating abdominal injury? A review and meta-analysis. AJR Am J Roentgenol. 2009;193(2):432-7.
242. Goonetilleke KS, Siriwardena AK**.** Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. Eur J Surg Oncol. 2007;33(3):266-70.
243. Gordon I, Barkovics M, Pindoria S, et al**.** Primary vesicoureteric reflux as a predictor of renal damage in children hospitalized with urinary tract infection: a systematic review and meta-analysis. J Am Soc Nephrol. 2003;14(3):739-44.
244. Gorelick MH, Shaw KN**.** Screening tests for urinary tract infection in children: A meta-analysis. Pediatrics. 1999;104(5):e54.

245. Goto M, Noguchi Y, Koyama H, et al**.** Diagnostic value of adenosine deaminase in tuberculous pleural effusion: a meta-analysis. Ann Clin Biochem. 2003;40(Pt 4):374-81.
246. Gottlieb RH, Widjaja J, Tian L, et al**.** Calf sonography for detecting deep venous thrombosis in symptomatic patients: experience and review of the literature. J Clin Ultrasound. 1999;27(8):415-20.
247. Gould MK, Kuschner WG, Rydzak CE, et al. Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: a meta-analysis. Ann Intern Med. 2003;139(11):879-92.
248. Gould MK, Maclean CC, Kuschner WG, et al**.** Accuracy of positron emission tomography for diagnosis of pulmonary nodules and mass lesions: a meta-analysis. JAMA. 2001;285(7):914-24.
249. Greco S, Girardi E, Masciangelo R, et al**.** Adenosine deaminase and interferon gamma measurements for the diagnosis of tuberculous pleurisy: a meta-analysis. Int J Tuberc Lung Dis. 2003;7(8):777-86.
250. Greco S, Rulli M, Girardi E, et al**.** Diagnostic accuracy of in-house PCR for pulmonary tuberculosis in smear-positive patients: meta-analysis and metaregression. J Clin Microbiol. 2009;47(3):569-76.
251. Gross SJ, Shulman LP, Tolley EA, et al. Isolated fetal choroid plexus cysts and trisomy 18: a review and meta-analysis. Am J Obstet Gynecol. 1995;172(1 Pt 1):83-7.
252. Gu P, Huang G, Chen Y,et al**.** Diagnostic utility of pleural fluid carcinoembryonic antigen and CYFRA 21-1 in patients with pleural effusion: a systematic review and meta-analysis. J Clin Lab Anal. 2007;21(6):398-405.
253. Gu P, Pan LL, Wu SQ, et al**.** CA 125, PET alone, PET-CT, CT and MRI in diagnosing recurrent ovarian carcinoma: a systematic review and meta-analysis. Eur J Radiol. 2009;71(1):164-74.
254. Gu P, Zhao YZ, Jiang LY, et al**.** Endobronchial ultrasound-guided transbronchial needle aspiration for staging of lung cancer: a systematic review and meta-analysis. Eur J Cancer. 2009;45(8):1389-96.
255. Gupta R, Dastane AM, McKenna R, Jr., et al**.** The predictive value of epidermal growth factor receptor tests in patients with pulmonary adenocarcinoma: review of current "best evidence" with meta-analysis. Hum Pathol. 2009;40(3):356-65.
256. Haase M, Bellomo R, Devarajan P, et al**.** Accuracy of neutrophil gelatinase-associated lipocalin (NGAL) in diagnosis and prognosis in acute kidney injury: a systematic review and meta-analysis. Am J Kidney Dis. 2009;54(6):1012-24.
257. Hajdinjak T**.** UroVysion FISH test for detecting urothelial cancers: meta-analysis of diagnostic accuracy and comparison with urinary cytology testing. Urol Oncol. 2008;26(6):646-51.
258. Halfon P, Munteanu M, Poynard T**.** FibroTest-ActiTest as a non-invasive marker of liver fibrosis. Gastroenterol Clin Biol. 2008;32(6 Suppl 1):22-39.
259. Hallan S, Asberg A**.** The accuracy of C-reactive protein in diagnosing acute appendicitis—a meta-analysis. Scand J Clin Lab Invest. 1997;57(5):373-80.
260. Halligan S, Altman DG, Taylor SA, et al. CT colonography in the detection of colorectal polyps and cancer: systematic review, meta-analysis, and proposed minimum data set for study level reporting. Radiology. 2005;237(3):893-904.
261. Hamon M, Biondi-Zoccai GG, Malagutti P, et al**.** Diagnostic performance of multislice spiral computed tomography of coronary arteries as compared with conventional invasive coronary angiography: a meta-analysis. J Am Coll Cardiol. 2006;48(9):1896-910.
262. Hamon M, Champ-Rigot L, Morello R, et al**.** Diagnostic accuracy of in-stent coronary restenosis detection with multislice spiral computed tomography: a meta-analysis. Eur Radiol. 2008;18(2):217-25.

263. Hamon M, Lepage O, Malagutti P, et al. Diagnostic performance of 16- and 64-section spiral CT for coronary artery bypass graft assessment: meta-analysis. Radiology. 2008;247(3):679-86.
264. Hamon M, Morello R, Riddell JW. Coronary arteries: diagnostic performance of 16- versus 64-section spiral CT compared with invasive coronary angiography—meta-analysis. Radiology. 2007;245(3):720-31.
265. Hancock MJ, Maher CG, Latimer J, et al. Systematic review of tests to identify the disc, SIJ or facet joint as the source of low back pain. Eur Spine J. 2007;16(10):1539-50.
266. Harvey P, Basuita A, Endersby D, et al. A systematic review of the diagnostic accuracy of prostate specific antigen. BMC Urol. 2009;9:14.
267. Harvey RT, Gefter WB, Hrung JM, et al. Accuracy of CT angiography versus pulmonary angiography in the diagnosis of acute pulmonary embolism: evaluation of the literature with summary ROC curve analysis. Acad Radiol. 2000;7(10):786-97.
268. Hatfield AS, Sanchez-Ramos L, Kaunitz AM. Sonographic cervical assessment to predict the success of labor induction: a systematic review with metaanalysis. Am J Obstet Gynecol. 2007;197(2):186-92.
269. Hayashino Y, Goto M, Noguchi Y, et al. Ventilation-perfusion scanning and helical CT in suspected pulmonary embolism: meta-analysis of diagnostic performance. Radiology. 2005;234(3):740-8.
270. Hegedus EJ, Cook C, Hasselblad V, et al. Physical examination tests for assessing a torn meniscus in the knee: a systematic review with meta-analysis. J Orthop Sports Phys Ther. 2007;37(9):541-50.
271. Hegedus EJ, Goode A, Campbell S, et al. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. Br J Sports Med. 2008;42(2):80-92; discussion
272. Heijenbrok-Kal MH, Fleischmann KE, Hunink MG. Stress echocardiography, stress single-photon-emission computed tomography and electron beam computed tomography for the assessment of coronary artery disease: a meta-analysis of diagnostic performance. Am Heart J. 2007;154(3):415-23.
273. Heijenbrok-Kal MH, Kock MC, Hunink MG. Lower extremity arterial disease: multidetector CT angiography meta-analysis. Radiology. 2007;245(2):433-9.
274. Heim SW, Schectman JM, Siadaty MS, et al. D-dimer testing for deep venous thrombosis: a metaanalysis. Clin Chem. 2004;50(7):1136-47.
275. Helbert M, Bodger S, Cavenagh J, et al. Optimising testing for phospholipid antibodies. J Clin Pathol. 2001;54(9):693-8.
276. Hendriks DJ, Kwee J, Mol BW, et al. Ultrasonography as a tool for the prediction of outcome in IVF patients: a comparative meta-analysis of ovarian volume and antral follicle count. Fertil Steril. 2007;87(4):764-75.
277. Hendriks DJ, Mol BW, Bancsi LF, et al. Antral follicle count in the prediction of poor ovarian response and pregnancy after in vitro fertilization: a meta-analysis and comparison with basal follicle-stimulating hormone level. Fertil Steril. 2005;83(2):291-301.
278. Henschke N, Maher CG, Refshauge KM. Screening for malignancy in low back pain patients: a systematic review. Eur Spine J. 2007;16(10):1673-9.
279. Hobby JL, Tom BD, Bearcroft PW, et al. Magnetic resonance imaging of the wrist: diagnostic performance statistics. Clin Radiol. 2001;56(1):50-7.
280. Hoffman RM, Clanon DL, Littenberg B, et al. Using the free-to-total prostate-specific antigen ratio to detect prostate cancer in men with nonspecific elevations of prostate-specific antigen levels. J Gen Intern Med. 2000;15(10):739-48.

281. Hofman PA, Nelemans P, Kemerink GJ, et al. Value of radiological diagnosis of skull fracture in the management of mild head injury: meta-analysis. J Neurol Neurosurg Psychiatry. 2000;68(4):416-22.
282. Hogg K, Brown G, Dunning J, et al. Diagnosis of pulmonary embolism with CT pulmonary angiography: a systematic review. Emerg Med J. 2006;23(3):172-8.
283. Holmes JF, Akkinepalli R. Computed tomography versus plain radiography to screen for cervical spine injury: a meta-analysis. J Trauma. 2005;58(5):902-5.
284. Holmes JF, Gladman A, Chang CH. Performance of abdominal ultrasonography in pediatric blunt trauma patients: a meta-analysis. J Pediatr Surg. 2007;42(9):1588-94.
285. Holroyd-Leduc JM, Tannenbaum C, Thorpe KE, et al. What type of urinary incontinence does this woman have? JAMA. 2008;299(12):1446-56.
286. Holty JE, Kuschner WG, Gould MK. Accuracy of transbronchial needle aspiration for mediastinal staging of non-small cell lung cancer: a meta-analysis. Thorax. 2005;60(11):949-55.
287. Honest H, Bachmann LM, Coomarasamy A, et al. Accuracy of cervical transvaginal sonography in predicting preterm birth: a systematic review. Ultrasound Obstet Gynecol. 2003;22(3):305-22.
288. Honest H, Bachmann LM, Gupta JK, et al. Accuracy of cervicovaginal fetal fibronectin test in predicting risk of spontaneous preterm birth: systematic review. BMJ. 2002;325(7359):301.
289. Honest H, Bachmann LM, Knox EM, et al. The accuracy of various tests for bacterial vaginosis in predicting preterm birth: a systematic review. BJOG. 2004;111(5):409-22.
290. Honest H, Bachmann LM, Ngai C, et al. The accuracy of maternal anthropometry measurements as predictor for spontaneous preterm birth—a systematic review. Eur J Obstet Gynecol Reprod Biol. 2005;119(1):11-20.
291. Honest H, Bachmann LM, Sengupta R, et al. Accuracy of absence of fetal breathing movements in predicting preterm birth: a systematic review. Ultrasound Obstet Gynecol. 2004;24(1):94-100.
292. Honest H, Sharma S, Khan KS. Rapid tests for group B Streptococcus colonization in laboring women: a systematic review. Pediatrics. 2006;117(4):1055-66.
293. Hoogendam A, Buntinx F, de Vet HC. The diagnostic value of digital rectal examination in primary care screening for prostate cancer: a meta-analysis. Fam Pract. 1999;16(6):621-6.
294. Horsthuis K, Bipat S, Bennink RJ, et al. Inflammatory bowel disease diagnosed with US, MR, scintigraphy, and CT: meta-analysis of prospective studies. Radiology. 2008;247(1):64-79.
295. Horsthuis K, Bipat S, Stokkers PC, et al. Magnetic resonance imaging for evaluation of disease activity in Crohn's disease: a systematic review. Eur Radiol. 2009;19(6):1450-60.
296. Houssami N, Ciatto S, Macaskill P, et al. Accuracy and surgical impact of magnetic resonance imaging in breast cancer staging: systematic review and meta-analysis in detection of multifocal and multicentric cancer. J Clin Oncol. 2008;26(19):3248-58.
297. Hovels AM, Heesakkers RA, Adang EM, et al. The diagnostic accuracy of CT and MRI in the staging of pelvic lymph nodes in patients with prostate cancer: a meta-analysis. Clin Radiol. 2008;63(4):387-95.
298. Hrung JM, Sonnad SS, Schwartz JS, et al. Accuracy of MR imaging in the work-up of suspicious breast lesions: a diagnostic meta-analysis. Acad Radiol. 1999;6(7):387-97.
299. Huang SH, Hwang D, Lockwood G, et al. Predictive value of tumor thickness for cervical lymph-node involvement in squamous cell carcinoma of the oral cavity: a meta-analysis of reported studies. Cancer. 2009;115(7):1489-97.

300. Huebner RH, Park KC, Shepherd JE, et al. A meta-analysis of the literature for whole-body FDG PET detection of recurrent colorectal cancer. J Nucl Med. 2000;41(7):1177-89.
301. Huicho L, Campos M, Rivera J, et al**.** Fecal screening tests in the approach to acute infectious diarrhea: a scientific overview. Pediatr Infect Dis J. 1996;15(6):486-94.
302. Huicho L, Campos-Sanchez M, Alamo C**.** Metaanalysis of urine screening tests for determining the risk of urinary tract infection in children. Pediatr Infect Dis J. 2002;21(1):1-11, 88.
303. Hurlbut TA, 3rd, Littenberg B**.** The diagnostic accuracy of rapid dipstick tests to predict urinary tract infection. Am J Clin Pathol. 1991;96(5):582-8.
304. Hurley JC**.** Concordance of endotoxemia with gram-negative bacteremia. A meta-analysis using receiver operating characteristic curves. Arch Pathol Lab Med. 2000;124(8):1157-64.
305. Hussain R, Buscombe JR**.** A meta-analysis of scintimammography: an evidence-based approach to its clinical utility. Nucl Med Commun. 2006;27(7):589-94.
306. Ie YL, Verdonschot EH**.** Performance of diagnostic systems in occlusal caries detection compared. Community Dent Oral Epidemiol. 1994;22(3):187-91.
307. Imran MB, Khan MA, Aslam MN, et al**.** Diagnosis of coronary artery disease by stress echocardiography and perfusion scintigraphy. J Coll Physicians Surg Pak. 2003;13(8):465-70.
308. Inaba Y, Bergmann SR**.** Diagnostic accuracy of beta-methyl-p-[123I]-iodophenyl-pentadecanoic acid (BMIPP) imaging: a meta-analysis. J Nucl Cardiol. 2008;15(3):345-52.
309. Ioannidis JP, Salem D, Chew PW, et al**.** Accuracy of imaging technologies in the diagnosis of acute cardiac ischemia in the emergency department: a meta-analysis. Ann Emerg Med. 2001;37(5):471-7.
310. Ioannidis JP, Salem D, Chew PW, et al**.** Accuracy and clinical effect of out-of-hospital electrocardiography in the diagnosis of acute cardiac ischemia: a meta-analysis. Ann Emerg Med. 2001;37(5):461-70.
311. Isasi CR, Lu P, Blaufox MD**.** A metaanalysis of 18F-2-deoxy-2-fluoro-D-glucose positron emission tomography in the staging and restaging of patients with lymphoma. Cancer. 2005;104(5):1066-74.
312. Isasi CR, Moadel RM, Blaufox MD**.** A meta-analysis of FDG-PET for the evaluation of breast cancer recurrence and metastases. Breast Cancer Res Treat. 2005;90(2):105-12.
313. Isles MG, McConkey C, Mehanna HM**.** A systematic review and meta-analysis of the role of positron emission tomography in the follow up of head and neck squamous cell carcinoma following radiotherapy or chemoradiotherapy. Clin Otolaryngol. 2008;33(3):210-22.
314. Jahromi AS, Cina CS, Liu Y, et al**.** Sensitivity and specificity of color duplex ultrasound measurement in the estimation of internal carotid artery stenosis: a systematic review and meta-analysis. J Vasc Surg. 2005;41(6):962-72.
315. Jain T, Soules MR, Collins JA**.** Comparison of basal follicle-stimulating hormone versus the clomiphene citrate challenge test for ovarian reserve screening. Fertil Steril. 2004;82(1):180-5.
316. James A, Matchar DB, Myers ER**.** Testing for von Willebrand disease in women with menorrhagia: a systematic review. Obstet Gynecol. 2004;104(2):381-8.
317. Janne d'Othee B, Siebert U, Cury R, et al**.** A systematic review on diagnostic accuracy of CT-based detection of significant coronary artery disease. Eur J Radiol. 2008;65(3):449-61.

318. Jellema P, van der Windt DA, Schellevis FG, et al. Systematic review: accuracy of symptom-based criteria for diagnosis of irritable bowel syndrome in primary care. Aliment Pharmacol Ther. 2009;30(7):695-706.
319. Jensen JE, Nielsen SH, Foged L, et al. The MICRAL test for diabetic microalbuminuria: predictive values as a function of prevalence. Scand J Clin Lab Invest. 1996;56(2):117-22.
320. Jiang J, Shi HZ, Liang QL, et al. Diagnostic value of interferon-gamma in tuberculous pleurisy: a metaanalysis. Chest. 2007;131(4):1133-41.
321. Jimenez D, Uresandi F, Otero R, et al. Troponin-based risk stratification of patients with acute nonmassive pulmonary embolism: systematic review and metaanalysis. Chest. 2009;136(4):974-82.
322. Jing JY, Huang TC, Cui W, et al. Should FEV1/FEV6 replace FEV1/FVC ratio to detect airway obstruction? A metaanalysis. Chest. 2009;135(4):991-8.
323. Jiyong J, Tiancha H, Wei C, et al. Diagnostic value of the soluble triggering receptor expressed on myeloid cells-1 in bacterial infection: a meta-analysis. Intensive Care Med. 2009;35(4):587-95.
324. Jones AE, Fiechtl JF, Brown MD, et al. Procalcitonin test in the diagnosis of bacteremia: a meta-analysis. Ann Emerg Med. 2007;50(1):34-41.
325. Jones CM, Athanasiou T. Is virtual bronchoscopy an efficient diagnostic tool for the thoracic surgeon? Ann Thorac Surg. 2005;79(1):365-74.
326. Jones CM, Athanasiou T, Dunne N, et al. Multi-detector computed tomography in coronary artery bypass graft assessment: a meta-analysis. Ann Thorac Surg. 2007;83(1):341-8.
327. Jones CM, Athanasiou T, Nair S, et al. Do technical parameters affect the diagnostic accuracy of virtual bronchoscopy in patients with suspected airways stenosis? Eur J Radiol. 2005;55(3):445-51.
328. Jones CM, Athanasiou T, Tekkis PP, et al. Does Doppler echography have a diagnostic role in patency assessment of internal thoracic artery grafts? Eur J Cardiothorac Surg. 2005;28(5):692-700.
329. Jorgensen K, Andersen TJ, Dam H. The diagnostic efficiency of the Rorschach Depression Index and the Schizophrenia Index: a review. Assessment. 2000;7(3):259-80.
330. Jorm AF. Methods of screening for dementia: a meta-analysis of studies comparing an informant questionnaire with a brief cognitive test. Alzheimer Dis Assoc Disord. 1997;11(3):158-62.
331. Joshi U, Raijmakers PG, Riphagen, II, et al. Attenuation-corrected vs. nonattenuation-corrected 2-deoxy-2-[F-18]fluoro-D-glucose-positron emission tomography in oncology: a systematic review. Mol Imaging Biol. 2007;9(3):99-105.
332. Kalantri S, Pai M, Pascopella L, et al. Bacteriophage- based tests for the detection of Mycobacterium tuberculosis in clinical specimens: a systematic review and meta-analysis. BMC Infect Dis. 2005;5:59.
333. Kallmes DF, Omary RA, Dix JE, et al. Specificity of MR angiography as a confirmatory test of carotid artery stenosis. AJNR Am J Neuroradiol. 1996;17(8):1501-6.
334. Kapoor A, Page S, Lavalley M, et al. Magnetic resonance imaging for diagnosing foot osteomyelitis: a meta-analysis. Arch Intern Med. 2007;167(2):125-32.
335. Karassa FB, Matsagas MI, Schmidt WA, et al. Meta-analysis: test performance of ultrasonography for giant-cell arteritis. Ann Intern Med. 2005;142(5):359-69.
336. Kardaun JW, Kardaun OJ. Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation. Methods Inf Med. 1990;29(1):12-22.
337. Karger R, Donner-Banzhoff N, Muller HH, et al. Diagnostic performance of the platelet function analyzer (PFA-100) for the detection of disorders of primary haemostasis in

patients with a bleeding history-a systematic review and meta-analysis. Platelets. 2007;18(4):249-60.
338. Kassai B, Boissel JP, Cucherat M, et al. A systematic review of the accuracy of ultrasound in the diagnosis of deep venous thrombosis in asymptomatic patients. Thromb Haemost. 2004;91(4):655-66.
339. Kearon C, Julian JA, Newman TE, et al. Noninvasive diagnosis of deep venous thrombosis. McMaster Diagnostic Imaging Practice Guidelines Initiative. Ann Intern Med. 1998;128(8):663-77.
340. Kellen M, Aronson S, Roizen MF, et al. Predictive and diagnostic tests of renal failure: a review. Anesth Analg. 1994;78(1):134-42.
341. Kelly S, Harris KM, Berry E, et al. A systematic review of the staging performance of endoscopic ultrasound in gastro-oesophageal carcinoma. Gut. 2001;49(4):534-9.
342. Kertai MD, Boersma E, Bax JJ, et al. A meta-analysis comparing the prognostic accuracy of six diagnostic tests for predicting perioperative cardiac risk in patients undergoing major vascular surgery. Heart. 2003;89(11):1327-34.
343. Khan S, Cloud GC, Kerry S, et al. Imaging of vertebral artery stenosis: a systematic review. J Neurol Neurosurg Psychiatry. 2007;78(11):1218-25.
344. Khunti K, Squire I, Abrams KR, et al. Accuracy of a 12-lead electrocardiogram in screening patients with suspected heart failure for open access echocardiography: a systematic review and meta-analysis. Eur J Heart Fail. 2004;6(5):571-6.
345. Kievit W, de Bruin JH, Adang EM, et al. Current clinical selection strategies for identification of hereditary non-polyposis colorectal cancer families are inadequate: a meta-analysis. Clin Genet. 2004;65(4):308-16.
346. Kim C, Kwok YS, Heagerty P, et al. Pharmacologic stress testing for coronary disease diagnosis: A meta-analysis. Am Heart J. 2001;142(6):934-44.
347. Kinkel K, Hricak H, Lu Y, et al. US characterization of ovarian masses: a meta-analysis. Radiology. 2000;217(3):803-11.
348. Kinkel K, Kaji Y, Yu KK, et al. Radiologic staging in patients with endometrial cancer: a meta-analysis. Radiology. 1999;212(3):711-8.
349. Kittler H, Pehamberger H, Wolff K, et al. Diagnostic accuracy of dermoscopy. Lancet Oncol. 2002;3(3):159-65.
350. Klassen TP, Rowe PC. Selecting diagnostic tests to identify febrile infants less than 3 months of age as being at low risk for serious bacterial infection: a scientific overview. J Pediatr. 1992;121(5 Pt 1):671-6.
351. Klompas M. Does this patient have an acute thoracic aortic dissection? JAMA. 2002;287(17):2262-72.
352. Koch H, Meerkerk GJ, Zaat JO, et al. Accuracy of carbohydrate-deficient transferrin in the detection of excessive alcohol consumption: a systematic review. Alcohol Alcohol. 2004;39(2):75-85.
353. Koelemay MJ, den Hartog D, Prins MH, et al. Diagnosis of arterial disease of the lower extremities with duplex ultrasonography. Br J Surg. 1996;83(3):404-9.
354. Koelemay MJ, Lijmer JG, Stoker J, et al. Magnetic resonance angiography for the evaluation of lower extremity arterial disease: a meta-analysis. JAMA. 2001;285(10):1338-45.
355. Koelemay MJ, Nederkoorn PJ, Reitsma JB, et al. Systematic review of computed tomographic angiography for assessment of carotid artery disease. Stroke. 2004;35(10):2306-12.
356. Koliopoulos G, Arbyn M, Martin-Hirsch P, et al. Diagnostic accuracy of human papillomavirus testing in primary cervical screening: a systematic review and meta-analysis of non-randomized studies. Gynecol Oncol. 2007;104(1):232-46.

357. Koopmans CM, van Pampus MG, Groen H, et al. Accuracy of serum uric acid as a predictive test for maternal complications in pre-eclampsia: bivariate meta-analysis and decision analysis. Eur J Obstet Gynecol Reprod Biol. 2009;146(1):8-14.
358. Koumans EH, Johnson RE, Knapp JS, et al. Laboratory testing for Neisseria gonorrhoeae by recently introduced nonculture tests: a performance review with clinical and public health considerations. Clin Infect Dis. 1998;27(5):1171-80.
359. Kowalski J, Tu XM, Jia G, et al. A comparative meta-analysis on the variability in test performance among FDA-licensed enzyme immunosorbent assays for HIV antibody testing. J Clin Epidemiol. 2001;54(5):448-61.
360. Kraag N, Thijs C, Knipschild P. Dyspepsia—how noisy are gallstones? A meta-analysis of epidemiologic studies of biliary pain, dyspeptic symptoms, and food intolerance. Scand J Gastroenterol. 1995;30(5):411-21.
361. Krug B, Crott R, Lonneux M, et al. Role of PET in the initial staging of cutaneous malignant melanoma: systematic review. Radiology. 2008;249(3):836-44.
362. Kumar Y, Gurusamy K, Pamecha V, et al. Tumor M2-pyruvate kinase as tumor marker in exocrine pancreatic cancer a meta-analysis. Pancreas. 2007;35(2):114-9.
363. Kumbhani DJ, Ingelmo CP, Schoenhagen P, et al. Meta-analysis of diagnostic efficacy of 64-slice computed tomography in the evaluation of coronary in-stent restenosis. Am J Cardiol. 2009;103(12):1675-81.
364. Kwee TC, Kwee RM. MR angiography in the follow-up of intracranial aneurysms treated with Guglielmi detachable coils: systematic review and meta-analysis. Neuroradiology. 2007;49(9):703-13.
365. Kwee TC, Kwee RM. Combined FDG-PET/CT for the detection of unknown primary tumors: systematic review and meta-analysis. Eur Radiol. 2009;19(3):731-44.
366. Kwee TC, Kwee RM, Alavi A. FDG-PET for diagnosing prosthetic joint infection: systematic review and metaanalysis. Eur J Nucl Med Mol Imaging. 2008;35(11):2122-32.
367. Kwok Y, Kim C, Grady D, et al. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol. 1999;83(5):660-6.
368. Kyzas PA, Evangelou E, Denaxa-Kyza D, et al. 18F-fluorodeoxyglucose positron emission tomography to evaluate cervical node metastases in patients with head and neck squamous cell carcinoma: a meta-analysis. J Natl Cancer Inst. 2008;100(10):712-20.
369. Lacasse Y, Wong E, Guyatt GH, et al. Transthoracic needle aspiration biopsy for the diagnosis of localised pulmonary lesions: a meta-analysis. Thorax. 1999;54(10):884-93.
370. Lahaye MJ, Engelen SM, Nelemans PJ, et al. Imaging for predicting the risk factors—the circumferential resection margin and nodal disease—of local recurrence in rectal cancer: a meta-analysis. Semin Ultrasound CT MR. 2005;26(4):259-68.
371. Lameris W, van Randen A, Bipat S, et al. Graded compression ultrasonography and computed tomography in acute colonic diverticulitis: meta-analysis of test accuracy. Eur Radiol. 2008;18(11):2498-511.
372. Latour-Perez J, Coves-Orts FJ, Abad-Terrado C, et al. Accuracy of B-type natriuretic peptide levels in the diagnosis of left ventricular dysfunction and heart failure: a systematic review. Eur J Heart Fail. 2006;8(4):390-9.
373. Lau J, Ioannidis JP, Balk EM, et al. Diagnosing acute cardiac ischemia in the emergency department: a systematic review of the accuracy and clinical effect of current technologies. Ann Emerg Med. 2001;37(5):453-60.
374. Lawrence WF, Liang W, Mandelblatt JS, et al. Serendipity in diagnostic imaging: magnetic resonance imaging of the breast. J Natl Cancer Inst. 1998;90(23):1792-800.
375. Leal YA, Flores LL, Garcia-Cortes LB, et al. Antibody-based detection tests for the diagnosis of Helicobacter pylori infection in children: a meta-analysis. PLoS One. 2008;3(11):e3751.

376. Lederle FA, Simel DL**.** The rational clinical examination. Does this patient have abdominal aortic aneurysm? JAMA. 1999;281(1):77-82.
377. Ledro-Cano D**.** Suspected choledocholithiasis: endoscopic ultrasound or magnetic resonance cholangio-pancreatography? A systematic review. Eur J Gastroenterol Hepatol. 2007;19(11):1007-11.
378. Lee A, Fan LT, Gin T, et al**.** A systematic review (meta-analysis) of the accuracy of the Mallampati tests to predict the difficult airway. Anesth Analg. 2006;102(6):1867-78.
379. Lee JB, Ahmad S, Gale CP**.** Detection of coins ingested by children using a handheld metal detector: a systematic review. Emerg Med J. 2005;22(12):839-44.
380. Lee R, Localio AR, Armstrong K, et al**.** A meta-analysis of the performance characteristics of the free prostate-specific antigen test. Urology. 2006;67(4):762-8.
381. Leitich H, Kaider A**.** Fetal fibronectin—how useful is it in the prediction of preterm birth? BJOG. 2003;110 Suppl 20:66-70.
382. Lensing AW, Hirsh J**.** 125I-fibrinogen leg scanning: reassessment of its role for the diagnosis of venous thrombosis in post-operative patients. Thromb Haemost. 1993;69(1):2-7.
383. LeSar CJ, Meier GH, DeMasi RJ, et al. The utility of color duplex ultrasonography in the diagnosis of temporal arteritis. J Vasc Surg. 2002;36(6):1154-60.
384. Lewis NR, Scott BB**.** Systematic review: the use of serology to exclude or diagnose coeliac disease (a comparison of the endomysial and tissue transglutaminase antibody tests). Aliment Pharmacol Ther. 2006;24(1):47-54.
385. Li J**.** Capnography alone is imperfect for endotracheal tube placement confirmation during emergency intubation. J Emerg Med. 2001;20(3):223-9.
386. Liang QL, Shi HZ, Qin XJ, et al**.** Diagnostic accuracy of tumour markers for malignant pleural effusion: a meta-analysis. Thorax. 2008;63(1):35-41.
387. Liang QL, Shi HZ, Wang K, et al**.** Diagnostic accuracy of adenosine deaminase in tuberculous pleurisy: a meta-analysis. Respir Med. 2008;102(5):744-54.
388. Liberman M, Sampalis F, Mulder DS, et al**.** Breast cancer diagnosis by scintimammography: a meta-analysis and review of the literature. Breast Cancer Res Treat. 2003;80(1):115-26.
389. Liedberg J, Panmekiate S, Petersson A, et al**.** Evidence-based evaluation of three imaging methods for the temporomandibular disc. Dentomaxillofac Radiol. 1996;25(5):234-41.
390. Ling DI, Flores LL, Riley LW, et al**.** Commercial nucleic-acid amplification tests for diagnosis of pulmonary tuberculosis in respiratory specimens: meta-analysis and meta-regression. PLoS One. 2008;3(2):e1536.
391. Ling DI, Zwerling AA, Pai M**.** GenoType MTBDR assays for the diagnosis of multidrug-resistant tuberculosis: a meta-analysis. Eur Respir J. 2008;32(5):1165-74.
392. Littenberg B, Mushlin AI**.** Technetium bone scanning in the diagnosis of osteomyelitis: a meta-analysis of test performance. Diagnostic Technology Assessment Consortium. J Gen Intern Med. 1992;7(2):158-64.
393. Liu J, Xu Y, Wang J**.** Ultrasonography, computed tomography and magnetic resonance imaging for diagnosis of ovarian carcinoma. Eur J Radiol. 2007;62(3):328-34.
394. Liu T, Xu W, Yan WL, et al**.** FDG-PET, CT, MRI for diagnosis of local residual or recurrent nasopharyngeal carcinoma, which one is the best? A systematic review. Radiother Oncol. 2007;85(3):327-35.
395. Locker T, Goodacre S, Sampson F, et al**.** Meta-analysis of plethysmography and rheography in the diagnosis of deep vein thrombosis. Emerg Med J. 2006;23(8):630-5.
396. Lord SJ, Lei W, Craft P, et al. A systematic review of the effectiveness of magnetic resonance imaging (MRI) as an addition to mammography and ultrasound in screening young women at high risk of breast cancer. Eur J Cancer. 2007;43(13):1905-17.

397. Lotan Y, Roehrborn CG. Sensitivity and specificity of commonly available bladder tumor markers versus cytology: results of a comprehensive literature review and meta-analyses. Urology. 2003;61(1):109-18; discussion 18.
398. Loy CT, Irwig LM, Katelaris PH, et al. Do commercial serological kits for Helicobacter pylori infection differ in accuracy? A meta-analysis. Am J Gastroenterol. 1996;91(6):1138-44.
399. Lyman GH, Kuderer NM. Gene expression profile assays as predictors of recurrence-free survival in early-stage breast cancer: a metaanalysis. Clin Breast Cancer. 2006;7(5):372-9.
400. Lysakowski C, Walder B, Costanza MC, et al. Transcranial Doppler versus angiography in patients with vasospasm due to a ruptured cerebral aneurysm: A systematic review. Stroke. 2001;32(10):2292-8.
401. Maas JW, Evers JL, ter Riet G, et al. Pregnancy rate following normal versus abnormal hysterosalpingography findings: a meta-analysis. Gynecol Obstet Invest. 1997;43(2):79-83.
402. MacDermid JC, Wessel J. Clinical diagnosis of carpal tunnel syndrome: a systematic review. J Hand Ther. 2004;17(2):309-19.
403. Mackenzie R, Palmer CR, Lomas DJ, et al. Magnetic resonance imaging of the knee: diagnostic performance studies. Clin Radiol. 1996;51(4):251-7.
404. Madhok V, Falk G, Rogers A, et al. The accuracy of symptoms, signs and diagnostic tests in the diagnosis of left ventricular dysfunction in primary care: a diagnostic accuracy systematic review. BMC Fam Pract. 2008;9:56.
405. Maenza RL, Seaberg D, D'Amico F. A meta-analysis of blunt cardiac trauma: ending myocardial confusion. Am J Emerg Med. 1996;14(3):237-41.
406. Maguire S, Pickerd N, Farewell D, et al. Which clinical features distinguish inflicted from non-inflicted brain injury? A systematic review. Arch Dis Child. 2009;94(11):860-7.
407. Maheshwari A, Gibreel A, Bhattacharya S, et al. Dynamic tests of ovarian reserve: a systematic review of diagnostic accuracy. Reprod Biomed Online. 2009;18(5):717-34.
408. Makrydimas G, Sotiriadis A, Ioannidis JP. Screening performance of first-trimester nuchal translucency for major cardiac defects: a meta-analysis. Am J Obstet Gynecol. 2003;189(5):1330-5.
409. Mallo RD, Salem L, Lalani T, et al. Computed tomography diagnosis of ischemia and complete obstruction in small bowel obstruction: a systematic review. J Gastrointest Surg. 2005;9(5):690-4.
410. Mann JJ, Currier D, Stanley B, et al. Can biological tests assist prediction of suicide in mood disorders? Int J Neuropsychopharmacol. 2006;9(4):465-74.
411. Mann RM, Hoogeveen YL, Blickman JG, et al. MRI compared to conventional diagnostic work-up in the detection and evaluation of invasive lobular carcinoma of the breast: a review of existing literature. Breast Cancer Res Treat. 2008;107(1):1-14.
412. Mantha S, Roizen MF, Barnard J, et al. Relative effectiveness of four preoperative tests for predicting adverse cardiac outcomes after vascular surgery: a meta-analysis. Anesth Analg. 1994;79(3):422-33.
413. Marik PE, Baram M, Vahid B. Does central venous pressure predict fluid responsiveness? A systematic review of the literature and the tale of seven mares. Chest. 2008;134(1):172-8.
414. Marik PE, Cavallazzi R, Vasu T, et al. Dynamic changes in arterial waveform derived variables and fluid responsiveness in mechanically ventilated patients: a systematic review of the literature. Crit Care Med. 2009;37(9):2642-7.
415. Marill KA. Serum D-dimer is a sensitive test for the detection of acute aortic dissection: a pooled meta-analysis. J Emerg Med. 2008;34(4):367-76.

416. Markert RJ, Walley ME, Guttman TG, et al. A pooled analysis of the Ottawa ankle rules used on adults in the ED. Am J Emerg Med. 1998;16(6):564-7.
417. Marmo R, Rotondano G, Piscopo R, et al. Meta-analysis: capsule enteroscopy vs. conventional modalities in diagnosis of small bowel diseases. Aliment Pharmacol Ther. 2005;22(7):595-604.
418. Martin A, Panaiotov S, Portaels F, et al. The nitrate reductase assay for the rapid detection of isoniazid and rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. J Antimicrob Chemother. 2008;62(1):56-64.
419. Martin A, Portaels F, Palomino JC. Colorimetric redox-indicator methods for the rapid detection of multidrug resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. J Antimicrob Chemother. 2007;59(2):175-83.
420. Martin JL, Williams KS, Sutton AJ, et al. Systematic review and meta-analysis of methods of diagnostic assessment for urinary incontinence. Neurourol Urodyn. 2006;25(7):674-83; discussion 84.
421. Marx A, Pewsner D, Egger M, et al. Meta-analysis: accuracy of rapid tests for malaria in travelers returning from endemic areas. Ann Intern Med. 2005;142(10):836-46.
422. Mase SR, Ramsay A, Ng V, et al. Yield of serial sputum specimen examinations in the diagnosis of pulmonary tuberculosis: a systematic review. Int J Tuberc Lung Dis. 2007;11(5):485-95.
423. McGee S, Abernethy WB, 3rd, Simel DL. The rational clinical examination. Is this patient hypovolemic? JAMA. 1999;281(11):1022-9.
424. Medeiros LR, Rosa DD, da Rosa MI, et al. Accuracy of ultrasonography with color Doppler in ovarian tumor: a systematic quantitative review. Int J Gynecol Cancer. 2009;19(7):1214-20.
425. Medeiros LR, Rosa DD, da Rosa MI, et al. Accuracy of CA 125 in the diagnosis of ovarian tumors: a quantitative systematic review. Eur J Obstet Gynecol Reprod Biol. 2009;142(2):99-105.
426. Medeiros LR, Rosa DD, Edelweiss MI, et al. Accuracy of frozen-section analysis in the diagnosis of ovarian tumors: a systematic quantitative review. Int J Gynecol Cancer. 2005;15(2):192-202.
427. Meijer AB, O YL, Geleijns J, et al. Meta-analysis of 40- and 64-MDCT angiography for assessing coronary artery stenosis. AJR Am J Roentgenol. 2008;191(6):1667-75.
428. Menke J. Diagnostic accuracy of contrast-enhanced MR angiography in severe carotid stenosis: meta-analysis with metaregression of different techniques. Eur Radiol. 2009;19(9):2204-16.
429. Menzies D, Pai M, Comstock G. Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research. Ann Intern Med. 2007;146(5):340-54.
430. Meredith SM, Sanchez-Ramos L, Kaunitz AM. Diagnostic accuracy of transvaginal sonography for the diagnosis of adenomyosis: systematic review and metaanalysis. Am J Obstet Gynecol. 2009;201(1):107 e1-6.
431. Merritt RM, Williams MF, James TH, et al. Detection of cervical metastasis. A meta-analysis comparing computed tomography with physical examination. Arch Otolaryngol Head Neck Surg. 1997;123(2):149-52.
432. Meserve BB, Cleland JA, Boucher TR. A meta-analysis examining clinical test utilities for assessing meniscal injury. Clin Rehabil. 2008;22(2):143-61.
433. Meserve BB, Cleland JA, Boucher TR. A meta-analysis examining clinical test utility for assessing superior labral anterior posterior lesions. Am J Sports Med. 2009;37(11):2252-8.

434. Met R, Bipat S, Legemate DA, et al**.** Diagnostic performance of computed tomography angiography in peripheral arterial disease: a systematic review and meta-analysis. JAMA. 2009;301(4):415-24.
435. Metz LD, Beattie M, Hom R, et al**.** The prognostic value of normal exercise myocardial perfusion imaging and exercise echocardiography: a meta-analysis. J Am Coll Cardiol. 2007;49(2):227-37.
436. Micames CG, McCrory DC, Pavey DA, et al**.** Endoscopic ultrasound-guided fine-needle aspiration for non-small cell lung cancer staging: A systematic review and metaanalysis. Chest. 2007;131(2):539-48.
437. Mijnhout GS, Hoekstra OS, van Tulder MW, et al**.** Systematic review of the diagnostic accuracy of (18)F-fluorodeoxyglucose positron emission tomography in melanoma patients. Cancer. 2001;91(8):1530-42.
438. Miltenburg DM, Nuchtern JG, Jaksic T, et al**.** Laparoscopic evaluation of the pediatric inguinal hernia—a meta-analysis. J Pediatr Surg. 1998;33(6):874-9.
439. Mirvis SE, Shanmuganathan K, Miller BH, et al**.** Traumatic aortic injury: diagnosis with contrast-enhanced thoracic CT—five-year experience at a major trauma center. Radiology. 1996;200(2):413-22.
440. Mistry K, Cable G**.** Meta-analysis of prostate-specific antigen and digital rectal examination as screening tests for prostate carcinoma. J Am Board Fam Pract. 2003;16(2):95-101.
441. Mitchell AJ**.** Are one or two simple questions sufficient to detect depression in cancer and palliative care? A Bayesian meta-analysis. Br J Cancer. 2008;98(12):1934-43.
442. Mitchell AJ**.** The clinical significance of subjective memory complaints in the diagnosis of mild cognitive impairment and dementia: a meta-analysis. Int J Geriatr Psychiatry. 2008;23(11):1191-202.
443. Mitchell AJ**.** CSF phosphorylated tau in the diagnosis and prognosis of mild cognitive impairment and Alzheimer's disease: a meta-analysis of 51 studies. J Neurol Neurosurg Psychiatry. 2009;80(9):966-75.
444. Mitchell AJ**.** A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. J Psychiatr Res. 2009;43(4):411-31.
445. Mitchell AJ, Coyne JC**.** Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. Br J Gen Pract. 2007;57(535):144-51.
446. Mitchell AJ, Vaze A, Rao S**.** Clinical diagnosis of depression in primary care: a meta-analysis. Lancet. 2009;374(9690):609-19.
447. Mitchell AM, Brown MD, Menown IB, et al**.** Novel protein markers of acute coronary syndrome complications in low-risk outpatients: a systematic review of potential use in the emergency department. Clin Chem. 2005;51(11):2005-12.
448. Mitchell MF, Cantor SB, Ramanujam N, et al**.** Fluorescence spectroscopy for diagnosis of squamous intraepithelial lesions of the cervix. Obstet Gynecol. 1999;93(3):462-70.
449. Mitchell MF, Schottenfeld D, Tortolero-Luna G, et al**.** Colposcopy for the diagnosis of squamous intraepithelial lesions: a meta-analysis. Obstet Gynecol. 1998;91(4):626-31.
450. Mofidi R, Suttie SA, Patil PV, et al**.** The value of procalcitonin at predicting the severity of acute pancreatitis and development of infected pancreatic necrosis: systematic review. Surgery. 2009;146(1):72-81.
451. Mol BW, Bayram N, Lijmer JG, et al. The performance of CA-125 measurement in the detection of endometriosis: a meta-analysis. Fertil Steril. 1998;70(6):1101-8.
452. Mol BW, Dijkman B, Wertheim P, et al**.** The accuracy of serum chlamydial antibodies in the diagnosis of tubal pathology: a meta-analysis. Fertil Steril. 1997;67(6):1031-7.

453. Mol BW, Lijmer JG, Ankum WM, et al**.** The accuracy of single serum progesterone measurement in the diagnosis of ectopic pregnancy: a meta-analysis. Hum Reprod. 1998;13(11):3220-7.
454. Mol BW, Meijer S, Yuppa S, et al. Sperm penetration assay in predicting successful in vitro fertilization. A meta-analysis. J Reprod Med. 1998;43(6):503-8.
455. Moles DR, Downer MC, Speight PM**.** Meta-analysis of measures of performance reported in oral cancer and precancer screening studies. Br Dent J. 2002;192(6):340-4; discussion 32.
456. Morgan M, Kalantri S, Flores L, et al**.** A commercial line probe assay for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a systematic review and meta-analysis. BMC Infect Dis. 2005;5:62.
457. Morisson P, Neves DD**.** Evaluation of adenosine deaminase in the diagnosis of pleural tuberculosis: a Brazilian meta-analysis. J Bras Pneumol. 2008;34(4):217-24.
458. Moroney JT, Bagiella E, Desmond DW, et al. Meta-analysis of the Hachinski Ischemic Score in pathologically verified dementias. Neurology. 1997;49(4):1096-105.
459. Morris RK, Cnossen JS, Langejans M, et al. Serum screening with Down's syndrome markers to predict pre-eclampsia and small for gestational age: systematic review and meta-analysis. BMC Pregnancy Childbirth. 2008;8:33.
460. Morris RK, Malin GL, Khan KS, et al**.** Antenatal ultrasound to predict postnatal renal function in congenital lower urinary tract obstruction: systematic review of test accuracy. BJOG. 2009;116(10):1290-9.
461. Morris RK, Quinlan-Jones E, Kilby MD, et al**.** Systematic review of accuracy of fetal urine analysis to predict poor postnatal renal function in cases of congenital urinary tract obstruction. Prenat Diagn. 2007;27(10):900-11.
462. Morselli-Labate AM, Pezzilli R**.** Usefulness of serum IgG4 in the diagnosis and follow up of autoimmune pancreatitis: A systematic literature review and meta-analysis. J Gastroenterol Hepatol. 2009;24(1):15-36.
463. Mos IC, Klok FA, Kroft LJ, et al**.** Safety of ruling out acute pulmonary embolism by normal computed tomography pulmonary angiography in patients with an indication for computed tomography: systematic review and meta-analysis. J Thromb Haemost. 2009;7(9):1491-8.
464. Mowatt G, Burr JM, Cook JA, et al. Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis. Invest Ophthalmol Vis Sci. 2008;49(12):5373-85.
465. Mowatt G, Cook JA, Hillis GS, et al. 64-Slice computed tomography angiography in the diagnosis and assessment of coronary artery disease: systematic review and meta-analysis. Heart. 2008;94(11):1386-93.
466. Muchow RD, Resnick DK, Abdel MP, et al**.** Magnetic resonance imaging (MRI) in the clearance of the cervical spine in blunt trauma: a meta-analysis. J Trauma. 2008;64(1):179-89.
467. Mulhall BP, Veerappan GR, Jackson JL**.** Meta-analysis: computed tomographic colonography. Ann Intern Med. 2005;142(8):635-50.
468. Munro W, Healy R**.** The validity and accuracy of clinical tests used to detect labral pathology of the shoulder—a systematic review. Man Ther. 2009;14(2):119-30.
469. Mushlin AI, Kouides RW, Shapiro DE**.** Estimating the accuracy of screening mammography: a meta-analysis. Am J Prev Med. 1998;14(2):143-53.
470. Nallamothu BK, Saint S, Bielak LF, et al. Electron-beam computed tomography in the diagnosis of coronary artery disease: a meta-analysis. Arch Intern Med. 2001;161(6):833-8.
471. Nandalur KR, Dwamena BA, Choudhri AF, et al**.** Diagnostic performance of stress cardiac magnetic resonance imaging in the detection of coronary artery disease: a meta-analysis. J Am Coll Cardiol. 2007;50(14):1343-53.

472. Nandalur KR, Dwamena BA, Choudhri AF, et al**.** Diagnostic performance of positron emission tomography in the detection of coronary artery disease: a meta-analysis. Acad Radiol. 2008;15(4):444-51.
473. Navare SM, Mather JF, Shaw LJ, et al**.** Comparison of risk stratification with pharmacologic and exercise stress myocardial perfusion imaging: a meta-analysis. J Nucl Cardiol. 2004;11(5):551-61.
474. Navarro JC, Lao AY, Sharma VK, et al**.** The accuracy of transcranial Doppler in the diagnosis of middle cerebral artery stenosis. Cerebrovasc Dis. 2007;23(5-6):325-30.
475. Nayak S, Olkin I, Liu H, et al. Meta-analysis: accuracy of quantitative ultrasound for identifying patients with osteoporosis. Ann Intern Med. 2006;144(11):832-41.
476. Nederkoorn PJ, van der Graaf Y, Hunink MG**.** Duplex ultrasound and magnetic resonance angiography compared with digital subtraction angiography in carotid artery stenosis: a systematic review. Stroke. 2003;34(5):1324-32.
477. Needham DM, Shufelt KA, Tomlinson G, et al**.** Troponin I and T levels in renal failure patients without acute coronary syndrome: a systematic review of the literature. Can J Cardiol. 2004;20(12):1212-8.
478. Nelemans PJ, Leiner T, de Vet HC, et al**.** Peripheral arterial disease: meta-analysis of the diagnostic performance of MR angiography. Radiology. 2000;217(1):105-14.
479. Ng PC, Dear PR**.** The predictive value of a normal ultrasound scan in the preterm baby—a meta-analysis. Acta Paediatr Scand. 1990;79(3):286-91.
480. Niemann T, Egelhof T, Bongartz G**.** Transthoracic sonography for the detection of pulmonary embolism—a meta-analysis. Ultraschall Med. 2009;30(2):150-6.
481. Niemann T, Kollmann T, Bongartz G**.** Diagnostic performance of low-dose CT for the detection of urolithiasis: a meta-analysis. AJR Am J Roentgenol. 2008;191(2):396-401.
482. Noble M, Bruening W, Uhl S, et al**.** Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. Arch Gynecol Obstet. 2009;279(6):881-90.
483. Noguchi Y, Nagata-Kobayashi S, Stahl JE, et al**.** A meta-analytic comparison of echocardiographic stressors. Int J Cardiovasc Imaging. 2005;21(2-3):189-207.
484. Norlund A, Axelsson S, Dahlen G, et al. Economic aspects of the detection of occlusal dentine caries. Acta Odontol Scand. 2009;67(1):38-43.
485. Nourbakhsh A, Grady JJ, Garges KJ**.** Percutaneous spine biopsy: a meta-analysis. J Bone Joint Surg Am. 2008;90(8):1722-5.
486. Numans ME, Lau J, de Wit NJ, et al**.** Short-term treatment with proton-pump inhibitors as a test for gastroesophageal reflux disease: a meta-analysis of diagnostic test characteristics. Ann Intern Med. 2004;140(7):518-27.
487. Ochoa ME, Marin Mdel C, Frutos-Vivar F, et al. Cuff-leak test for the diagnosis of upper airway obstruction in adults: a systematic review and meta-analysis. Intensive Care Med. 2009;35(7):1171-9.
488. Oei EH, Nikken JJ, Verstijnen AC, et al**.** MR imaging of the menisci and cruciate ligaments: a systematic review. Radiology. 2003;226(3):837-48.
489. Oei SG, Helmerhorst FM, Keirse MJ**.** When is the post-coital test normal? A critical appraisal. Hum Reprod. 1995;10(7):1711-4.
490. Offringa M, Benbassat J**.** The value of urinary red cell shape in the diagnosis of glomerular and post-glomerular haematuria. A meta-analysis. Postgrad Med J. 1992;68(802):648-54.
491. Ogilvie GS, Patrick DM, Schulzer M, et al. Diagnostic accuracy of self collected vaginal specimens for human papillomavirus compared to clinician collected human papillomavirus specimens: a meta-analysis. Sex Transm Infect. 2005;81(3):207-12.

492. Ola B, Afnan M, Papaioannou S, et al**.** Accuracy of sperm-cervical mucus penetration tests in evaluating sperm motility in semen: a systematic quantitative review. Hum Reprod. 2003;18(5):1037-46.
493. Olaniyan OB**.** Validity of colposcopy in the diagnosis of early cervical neoplasia—a review. Afr J Reprod Health. 2002;6(3):59-69.
494. Olatidoye AG, Wu AH, Feng YJ, et al**.** Prognostic role of troponin T versus troponin I in unstable angina pectoris for cardiac events with meta-analysis comparing published studies. Am J Cardiol. 1998;81(12):1405-10.
495. Orlando LA, Kulasingam SL, Matchar DB**.** Meta-analysis: the detection of pancreatic malignancy with positron emission tomography. Aliment Pharmacol Ther. 2004;20(10):1063-70.
496. Orr RK, Porter D, Hartman D**.** Ultrasonography to evaluate adults for appendicitis: decision making based on meta-analysis and probabilistic reasoning. Acad Emerg Med. 1995;2(7):644-50.
497. Owens DK, Holodniy M, Garber AM, et al. Polymerase chain reaction for the diagnosis of HIV infection in adults. A meta-analysis with recommendations for clinical practice and study design. Ann Intern Med. 1996;124(9):803-15.
498. Owens DK, Holodniy M, McDonald TW, et al**.** A meta-analytic evaluation of the polymerase chain reaction for the diagnosis of HIV infection in infants. JAMA. 1996;275(17):1342-8.
499. Pai M, Flores LL, Hubbard A, et al. Nucleic acid amplification tests in the diagnosis of tuberculous pleuritis: a systematic review and meta-analysis. BMC Infect Dis. 2004;4:6.
500. Pai M, Flores LL, Pai N, et al. Diagnostic accuracy of nucleic acid amplification tests for tuberculous meningitis: a systematic review and meta-analysis. Lancet Infect Dis. 2003;3(10):633-43.
501. Pai M, Kalantri S, Pascopella L, et al**.** Bacteriophage-based assays for the rapid detection of rifampicin resistance in Mycobacterium tuberculosis: a meta-analysis. J Infect. 2005;51(3):175-87.
502. Pai M, Zwerling A, Menzies D**.** Systematic review: T-cell-based assays for the diagnosis of latent tuberculosis infection: an update. Ann Intern Med. 2008;149(3):177-84.
503. Pai NP, Tulsky JP, Cohan D, et al**.** Rapid point-of-care HIV testing in pregnant women: a systematic review and meta-analysis. Trop Med Int Health. 2007;12(2):162-73.
504. Pakos EE, Fotopoulos AD, Ioannidis JP**.** 18F-FDG PET for evaluation of bone marrow infiltration in staging of lymphoma: a meta-analysis. J Nucl Med. 2005;46(6):958-63.
505. Pakos EE, Ioannidis JP**.** The association of P-glycoprotein with response to chemotherapy and clinical outcome in patients with osteosarcoma. A meta-analysis. Cancer. 2003;98(3):581-9.
506. Pakos EE, Koumoulis HD, Fotopoulos AD, et al**.** Osteomyelitis: antigranulocyte scintigraphy with 99mTC radiolabeled monoclonal antibodies for diagnosis— meta-analysis. Radiology. 2007;245(3):732-41.
507. Pakos EE, Kyzas PA, Ioannidis JP**.** Prognostic significance of TP53 tumor suppressor gene expression and mutations in human osteosarcoma: a meta-analysis. Clin Cancer Res. 2004;10(18 Pt 1):6208-14.
508. Pakos EE, Trikalinos TA, Fotopoulos AD, et al**.** Prosthesis infection: diagnosis after total joint arthroplasty with antigranulocyte scintigraphy with 99mTc-labeled monoclonal antibodies—a meta-analysis. Radiology. 2007;242(1):101-8.
509. Paleri V, Rees G, Arullendran P, et al**.** Sentinel node biopsy in squamous cell cancer of the oral cavity and oral pharynx: a diagnostic meta-analysis. Head Neck. 2005;27(9):739-47.
510. Palomaki GE, Neveux LM, Haddow JE**.** Can reliable Down's syndrome detection rates be determined from prenatal screening intervention trials? J Med Screen. 1996;3(1):12-7.

511. Parkes J, Guha IN, Roderick P, et al**.** Performance of serum marker panels for liver fibrosis in chronic hepatitis C. J Hepatol. 2006;44(3):462-74.
512. Patel SR, Wiese W, Patel SC, et al**.** Systematic review of diagnostic tests for vaginal trichomoniasis. Infect Dis Obstet Gynecol. 2000;8(5-6):248-57.
513. Patrick DL, Cheadle A, Thompson DC, et al**.** The validity of self-reported smoking: a review and meta-analysis. Am J Public Health. 1994;84(7):1086-93.
514. Patwardhan MB, McCrory DC, Matchar DB, et al**.** Alzheimer disease: operating characteristics of PET—a meta-analysis. Radiology. 2004;231(1):73-80.
515. Paulson WD, Ram SJ, Birk CG, et al**.** Does blood flow accurately predict thrombosis or failure of hemodialysis synthetic grafts? A meta-analysis. Am J Kidney Dis. 1999;34(3):478-85.
516. Peacock F, Morris DL, Anwaruddin S, et al. Meta-analysis of ischemia-modified albumin to rule out acute coronary syndromes in the emergency department. Am Heart J. 2006;152(2):253-62.
517. Peng Y, Wang HH**.** A meta-analysis of comparing fine-needle aspiration and frozen section for evaluating thyroid nodules. Diagn Cytopathol. 2008;36(12):916-20.
518. Peters NH, Borel Rinkes IH, Zuithoff NP, et al**.** Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology. 2008;246(1):116-24.
519. Petersen JR, Smith E, Okorodudu AO, et al**.** Comparison of four methods (L/S ratio, TDx FLM, lamellar bodies, PG) for fetal lung maturity using meta-analysis. Clin Lab Manage Rev. 1996;10(2):169-75.
520. Pfeiffer CD, Fine JP, Safdar N**.** Diagnosis of invasive aspergillosis using a galactomannan assay: a meta-analysis. Clin Infect Dis. 2006;42(10):1417-27.
521. Phillips KA**.** The use of meta-analysis in technology assessment: a meta-analysis of the enzyme immunosorbent assay human immunodeficiency virus antibody test. J Clin Epidemiol. 1991;44(9):925-31.
522. Picano E, Bedetti G, Varga A, et al**.** The comparable diagnostic accuracies of dobutamine-stress and dipyridamole-stress echocardiographies: a meta-analysis. Coron Artery Dis. 2000;11(2):151-9.
523. Picano E, Molinaro S, Pasanisi E**.** The diagnostic accuracy of pharmacological stress echocardiography for the assessment of coronary artery disease: a meta-analysis. Cardiovasc Ultrasound. 2008;6:30.
524. Pirozzo S, Papinczak T, Glasziou P**.** Whispered voice test for screening for hearing impairment in adults and children: systematic review. BMJ. 2003;327(7421):967.
525. Plevova P, Krepelova A, Papezova M, et al. Immunohistochemical detection of the hMLH1 and hMSH2 proteins in hereditary non-polyposis colon cancer and sporadic colon cancer. Neoplasma. 2004;51(4):275-84.
526. Plotkin M, Hautzel H, Krause BJ, et al. Implication of 2-18fluor-2-deoxyglucose positron emission tomography in the follow-up of Hurthle cell thyroid cancer. Thyroid. 2002;12(2):155-61.
527. Pretlove SJ, Fox CE, Khan KS, et al**.** Noninvasive methods of detecting fetal anaemia: a systematic review and meta-analysis. BJOG. 2009;116(12):1558-67.
528. Price CP, Newall RG, Boyd JC**.** Use of protein:creatinine ratio measurements on random urine samples for prediction of significant proteinuria: a systematic review. Clin Chem. 2005;51(9):1577-86.
529. Puli SR, Batapati Krishna Reddy J, Bechtold ML, et al**.** How good is endoscopic ultrasound for TNM staging of gastric cancers? A meta-analysis and systematic review. World J Gastroenterol. 2008;14(25):4011-9.
530. Puli SR, Batapati Krishna Reddy J, Bechtold ML, et al. Endoscopic ultrasound: it's accuracy in evaluating mediastinal lymphadenopathy? A meta-analysis and systematic review. World J Gastroenterol. 2008;14(19):3028-37.

531. Puli SR, Bechtold ML, Reddy JB, et al. How good is endoscopic ultrasound in differentiating various T stages of rectal cancer? Meta-analysis and systematic review. Ann Surg Oncol. 2009;16(2):254-65.
532. Puli SR, Reddy JB, Bechtold ML, et al. Staging accuracy of esophageal cancer by endoscopic ultrasound: a meta-analysis and systematic review. World J Gastroenterol. 2008;14(10):1479-90.
533. Puli SR, Reddy JB, Bechtold ML, et al. Accuracy of endoscopic ultrasound in the diagnosis of distal and celiac axis lymph node metastasis in esophageal cancer: a meta-analysis and systematic review. Dig Dis Sci. 2008;53(9):2405-14.
534. Puli SR, Reddy JB, Bechtold ML, et al. Accuracy of endoscopic ultrasound to diagnose nodal invasion by rectal cancers: a meta-analysis and systematic review. Ann Surg Oncol. 2009;16(5):1255-65.
535. Puli SR, Singh S, Hagedorn CH, et al. Diagnostic accuracy of EUS for vascular invasion in pancreatic and periampullary cancers: a meta-analysis and systematic review. Gastrointest Endosc. 2007;65(6):788-97.
536. Purkayastha S, Athanasiou T, Tekkis PP, et al. Magnetic resonance colonography vs computed tomography colonography for the diagnosis of colorectal cancer: an indirect comparison. Colorectal Dis. 2007;9(2):100-11.
537. Purkayastha S, Chow A, Athanasiou T, et al. Does serum procalcitonin have a role in evaluating the severity of acute pancreatitis? A question revisited. World J Surg. 2006;30(9):1713-21.
538. Purkayastha S, Tekkis PP, Athanasiou T, et al. Magnetic resonance colonography versus colonoscopy as a diagnostic investigation for colorectal cancer: a meta-analysis. Clin Radiol. 2005;60(9):980-9.
539. Purkayastha S, Tekkis PP, Athanasiou T, et al. Diagnostic precision of magnetic resonance imaging for preoperative prediction of the circumferential margin involvement in patients with rectal cancer. Colorectal Dis. 2007;9(5):402-11.
540. Quaglino P, Savoia P, Fierro MT, et al. Clinical significance of sequential tyrosinase expression in the peripheral blood of disease-free melanoma patients: a review of literature data. Melanoma Res. 2004;14(2):S17-9.
541. Quiroz R, Kucher N, Zou KH, et al. Clinical validity of a negative computed tomography scan in patients with suspected pulmonary embolism: a systematic review. JAMA. 2005;293(16):2012-7.
542. Raijmakers PG, Paul MA, Lips P. Sentinel node detection in patients with thyroid carcinoma: a meta-analysis. World J Surg. 2008;32(9):1961-7.
543. Rajpara SM, Botello AP, Townend J, et al. Systematic review of dermoscopy and digital dermoscopy/ artificial intelligence for the diagnosis of melanoma. Br J Dermatol. 2009;161(3):591-604.
544. Rao JK, Weinberger M, Oddone EZ, et al. The role of antineutrophil cytoplasmic antibody (c-ANCA) testing in the diagnosis of Wegener granulomatosis. A literature review and meta-analysis. Ann Intern Med. 1995;123(12):925-32.
545. Rasiah SV, Publicover M, Ewer AK, et al. A systematic review of the accuracy of first-trimester ultrasound examination for detecting major congenital heart disease. Ultrasound Obstet Gynecol. 2006;28(1):110-6.
546. Reed JF, 3rd. Meta-analysis of the reliability of noninvasive carotid studies. Biomed Instrum Technol. 1991;25(6):465-71.
547. Reed WW, Byrd GS, Gates RH, Jr., et al. Sputum gram's stain in community-acquired pneumococcal pneumonia. A meta-analysis. West J Med. 1996;165(4):197-204.
548. Reese GE, Constantinides VA, Simillis C, et al. Diagnostic precision of anti-Saccharomyces cerevisiae antibodies and perinuclear antineutrophil cytoplasmic antibodies in inflammatory bowel disease. Am J Gastroenterol. 2006;101(10):2410-22.

549. Revah A, Hannah ME, Sue-A-Quan AK**.** Fetal fibronectin as a predictor of preterm birth: an overview. Am J Perinatol. 1998;15(11):613-21.
550. Reznek L**.** The Rey 15-item memory test for malingering: a meta-analysis. Brain Inj. 2005;19(7):539-43.
551. Riquelme A, Calvo M, Salech F, et al. Value of adenosine deaminase (ADA) in ascitic fluid for the diagnosis of tuberculous peritonitis: a meta-analysis. J Clin Gastroenterol. 2006;40(8):705-10.
552. Roach ES**.** Tuberous sclerosis: function follows form. J Child Neurol. 1997;12(2):75-6.
553. Roberts AP, Childs SM, Rubin G, et al**.** Tests for Helicobacter pylori infection: a critical appraisal from primary care. Fam Pract. 2000;17 Suppl 2:S12-20.
554. Roddam AW, Duffy MJ, Hamdy FC, et al. Use of prostate-specific antigen (PSA) isoforms for the detection of prostate cancer in men with a PSA level of 2-10 ng/ml: systematic review and meta-analysis. Eur Urol. 2005;48(3):386-99; discussion 98-9.
555. Rodgers MA, Hempel S, Aho T, et al**.** Diagnostic tests used in the investigation of adult haematuria: A systematic review. BJU Int. 2006;98(6):1154-60.
556. Romagnuolo J, Bardou M, Rahme E, et al**.** Magnetic resonance cholangiopancreatography: a meta-analysis of test performance in suspected biliary disease. Ann Intern Med. 2003;139(7):547-57.
557. Roos JF, Doust J, Tett SE, et al**.** Diagnostic accuracy of cystatin C compared to serum creatinine for the estimation of renal dysfunction in adults and children—a meta-analysis. Clin Biochem. 2007;40(5-6):383-91.
558. Rosa MI, Medeiros LR, Bozzetti MC, et al. Accuracy of telomerase in cervical lesions: a systematic review. Int J Gynecol Cancer. 2007;17(6):1205-14.
559. Rosado B, Menzies S, Harbauer A, et al. Accuracy of computer diagnosis of melanoma: a quantitative meta-analysis. Arch Dermatol. 2003;139(3):361-7; discussion 6.
560. Rosenberg D, Cretin S**.** Use of meta-analysis to evaluate tolonium chloride in oral cancer screening. Oral Surg Oral Med Oral Pathol. 1989;67(5):621-7.
561. Rosman AS, Korsten MA**.** Meta-analysis comparing CT colonography, air contrast barium enema, and colonoscopy. Am J Med. 2007;120(3):203-10 e4.
562. Ross SD, Sheinhait IA, Harrison KJ, et al. Systematic review and meta-analysis of the literature regarding the diagnosis of sleep apnea. Sleep. 2000;23(4):519-32.
563. Rostom A, Dube C, Cranney A, et al. The diagnostic accuracy of serologic tests for celiac disease: a systematic review. Gastroenterology. 2005;128(4 Suppl 1):S38-46.
564. Roy PM, Colombet I, Durieux P, et al**.** Systematic review and meta-analysis of strategies for the diagnosis of suspected pulmonary embolism. BMJ. 2005;331(7511):259.
565. Rud B, Hilden J, Hyldstrup L, et al**.** Performance of the Osteoporosis Self-Assessment Tool in ruling out low bone mineral density in postmenopausal women: a systematic review. Osteoporos Int. 2007;18(9):1177-87.
566. Rud B, Hilden J, Hyldstrup L, et al**.** The Osteoporosis Self-Assessment Tool versus alternative tests for selecting postmenopausal women for bone mineral density assessment: a comparative systematic review of accuracy. Osteoporos Int. 2009;20(4):599-607.
567. Sadowski DC, Rabeneck L**.** Gastric ulcers at endoscopy: brush, biopsy, or both? Am J Gastroenterol. 1997;92(4):608-13.
568. Safdar N, Fine JP, Maki DG**.** Meta-analysis: methods for diagnosing intravascular device-related bloodstream infection. Ann Intern Med. 2005;142(6):451-66.
569. Safriel Y, Zinn H**.** CT pulmonary angiography in the detection of pulmonary emboli: a meta-analysis of sensitivities and specificities. Clin Imaging. 2002;26(2):101-5.
570. Sampson FC, Goodacre SW, Thomas SM, et al**.** The accuracy of MRI in diagnosis of suspected deep vein thrombosis: systematic review and meta-analysis. Eur Radiol. 2007;17(1):175-81.

571. Samson DJ, Flamm CR, Pisano ED, et al**.** Should FDG PET be used to decide whether a patient with an abnormal mammogram or breast finding at physical examination should undergo biopsy? Acad Radiol. 2002;9(7):773-83.
572. Sanchez-Ramos L, Delke I, Zamora J, et al**.** Fetal fibronectin as a short-term predictor of preterm birth in symptomatic patients: a meta-analysis. Obstet Gynecol. 2009;114(3):631-40.
573. Sanders S, Barnett A, Correa-Velez I, et al**.** Systematic review of the diagnostic accuracy of C-reactive protein to detect bacterial infection in nonhospitalized infants and children with fever. J Pediatr. 2008;153(4):570-4.
574. Sarmiento OL, Weigle KA, Alexander J, et al**.** Assessment by meta-analysis of PCR for diagnosis of smear-negative pulmonary tuberculosis. J Clin Microbiol. 2003;41(7):3233-40.
575. Sarwar A, Shaw LJ, Shapiro MD, et al. Diagnostic and prognostic value of absence of coronary artery calcification. JACC Cardiovasc Imaging. 2009;2(6):675-88.
576. Sauerland S, Bouillon B, Rixen D, et al**.** The reliability of clinical examination in detecting pelvic fractures in blunt trauma patients: a meta-analysis. Arch Orthop Trauma Surg. 2004;124(2):123-8.
577. Scheidler J, Hricak H, Yu KK, et al**.** Radiological evaluation of lymph node metastases in patients with cervical cancer. A meta-analysis. JAMA. 1997;278(13):1096-101.
578. Schinkel AF, Bax JJ, Geleijnse ML, et al. Noninvasive evaluation of ischaemic heart disease: myocardial perfusion imaging or stress echocardiography? Eur Heart J. 2003;24(9):789-800.
579. Scholten RJ, Deville WL, Opstelten W, et al**.** The accuracy of physical diagnostic tests for assessing meniscal lesions of the knee: a meta-analysis. J Fam Pract. 2001;50(11):938-44.
580. Scholten RJ, Opstelten W, van der Plas CG, et al**.** Accuracy of physical diagnostic tests for assessing ruptures of the anterior cruciate ligament: a meta-analysis. J Fam Pract. 2003;52(9):689-94.
581. Schreiber G, McCrory DC**.** Performance characteristics of different modalities for diagnosis of suspected lung cancer: summary of published evidence. Chest. 2003;123(1 Suppl):115S-28S.
582. Schuijf JD, Bax JJ, Shaw LJ, et al. Meta-analysis of comparative diagnostic performance of magnetic resonance imaging and multislice computed tomography for noninvasive coronary angiography. Am Heart J. 2006;151(2):404-11.
583. Schwimmer J, Essner R, Patel A, et al. A review of the literature for whole-body FDG PET in the management of patients with melanoma. Q J Nucl Med. 2000;44(2):153-67.
584. Scouller K, Conigrave KM, Macaskill P, et al**.** Should we use carbohydrate-deficient transferrin instead of gamma-glutamyltransferase for detecting problem drinkers? A systematic review and metaanalysis. Clin Chem. 2000;46(12):1894-902.
585. Selman TJ, Luesley DM, Acheson N, et al**.** A systematic review of the accuracy of diagnostic tests for inguinal lymph node status in vulvar cancer. Gynecol Oncol. 2005;99(1):206-14.
586. Selman TJ, Mann C, Zamora J, et al**.** Diagnostic accuracy of tests for lymph node status in primary cervical cancer: a systematic review and meta-analysis. CMAJ. 2008;178(7):855-62.
587. Selman TJ, Mann CH, Zamora J, et al**.** A systematic review of tests for lymph node status in primary endometrial cancer. BMC Womens Health. 2008;8:8.
588. Shafiq N, Malhotra S, Bhasin DK, et al**.** Estimating the diagnostic accuracy of procalcitonin as a marker of the severity of acute pancreatitis: a meta-analytic approach. JOP. 2005;6(3):231-7.

589. Shaheen AA, Myers RP**.** Diagnostic accuracy of the aspartate aminotransferase-to-platelet ratio index for the prediction of hepatitis C-related fibrosis: a systematic review. Hepatology. 2007;46(3):912-21.
590. Shaheen AA, Myers RP**.** Systematic review and meta-analysis of the diagnostic accuracy of fibrosis marker panels in patients with HIV/hepatitis C coinfection. HIV Clin Trials. 2008;9(1):43-51.
591. Shaheen AA, Wan AF, Myers RP**.** FibroTest and FibroScan for the prediction of hepatitis C-related fibrosis: a systematic review of diagnostic test accuracy. Am J Gastroenterol. 2007;102(11):2589-600.
592. Shaikh N, Morone NE, Lopez J, et al. Does this child have a urinary tract infection? JAMA. 2007;298(24):2895-904.
593. Sharpless KE, O'Sullivan DM, Schnatz PF**.** The utility of human papillomavirus testing in the management of atypical glandular cells on cytology. J Low Genit Tract Dis. 2009;13(2):72-8.
594. Shaw LJ, Peterson ED, Kesler K, et al**.** A metaanalysis of predischarge risk stratification after acute myocardial infarction with stress electrocardiographic, myocardial perfusion, and ventricular function imaging. Am J Cardiol. 1996;78(12):1327-37.
595. Shea JA, Berlin JA, Escarce JJ, et al. Revised estimates of diagnostic test sensitivity and specificity in suspected biliary tract disease. Arch Intern Med. 1994;154(22):2573-81.
596. Shi HZ, Liang QL, Jiang J, et al**.** Diagnostic value of carcinoembryonic antigen in malignant pleural effusion: a meta-analysis. Respirology. 2008;13(4):518-27.
597. Shie P, Cardarelli R, Brandon D, et al**.** Meta-analysis: comparison of F-18 Fluorodeoxyglucose-positron emission tomography and bone scintigraphy in the detection of bone metastases in patients with breast cancer. Clin Nucl Med. 2008;33(2):97-101.
598. Shiga T, Wajima Z, Apfel CC, et al**.** Diagnostic accuracy of transesophageal echocardiography, helical computed tomography, and magnetic resonance imaging for suspected thoracic aortic dissection: systematic review and meta-analysis. Arch Intern Med. 2006;166(13):1350-6.
599. Shiga T, Wajima Z, Inoue T, et al**.** Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. Anesthesiology. 2005;103(2):429-37.
600. Shimada T, Noguchi Y, Jackson JL, et al. Systematic review and metaanalysis: urinary antigen tests for Legionellosis. Chest. 2009;136(6):1576-85.
601. Siegman-Igra Y, Anglim AM, Shapiro DE, et al**.** Diagnosis of vascular catheter-related bloodstream infection: a meta-analysis. J Clin Microbiol. 1997;35(4):928-36.
602. Silvestri GA, Littenberg B, Colice GL**.** The clinical evaluation for detecting metastatic lung cancer. A meta-analysis. Am J Respir Crit Care Med. 1995;152(1):225-30.
603. Simon L, Gauvin F, Amre DK, et al**.** Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. Clin Infect Dis. 2004;39(2):206-17.
604. Singal A, Volk ML, Waljee A, et al. Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. Aliment Pharmacol Ther. 2009;30(1):37-47.
605. Sinuff T, Adhikari NK, Cook DJ, et al. Mortality predictions in the intensive care unit: comparing physicians with scoring systems. Crit Care Med. 2006;34(3):878-85.
606. Skupski DW, Rosenberg CR, Eglinton GS**.** Intrapartum fetal stimulation tests: a meta-analysis. Obstet Gynecol. 2002;99(1):129-34.
607. Sloka JS, Hollett PD, Mathews M**.** A quantitative review of the use of FDG-PET in the axillary staging of breast cancer. Med Sci Monit. 2007;13(3):RA37-46.

608. Smart SC, Sagar KB. Diagnostic and prognostic use of stress echocardiography in stable patients. Echocardiography. 2000;17(5):465-77.
609. Smith-Bindman R, Hosmer W, Feldstein VA, et al. Second-trimester ultrasound to detect fetuses with Down syndrome: a meta-analysis. JAMA. 2001;285(8):1044-55.
610. Smith-Bindman R, Kerlikowske K, Feldstein VA, et al. Endovaginal ultrasound to exclude endometrial cancer and other endometrial abnormalities. JAMA. 1998;280(17):1510-7.
611. Sodeck G, Domanovits H, Schillinger M, et al. D-dimer in ruling out acute aortic dissection: a systematic review and prospective cohort study. Eur Heart J. 2007;28(24):3067-75.
612. Solomon DH, Simel DL, Bates DW, et al. The rational clinical examination. Does this patient have a torn meniscus or ligament of the knee? Value of the physical examination. JAMA. 2001;286(13):1610-20.
613. Song IH, Carrasco-Fernandez J, Rudwaleit M, et al. The diagnostic value of scintigraphy in assessing sacroiliitis in ankylosing spondylitis: a systematic literature research. Ann Rheum Dis. 2008;67(11):1535-40.
614. Song JM, Kim CB, Chung HC, et al. Prostate-specific antigen, digital rectal examination and transrectal ultrasonography: a meta-analysis for this diagnostic triad of prostate cancer in symptomatic korean men. Yonsei Med J. 2005;46(3):414-24.
615. Sonnad SS, Langlotz CP, Schwartz JS. Accuracy of MR imaging for staging prostate cancer: a meta-analysis to examine the effect of technologic change. Acad Radiol. 2001;8(2):149-57.
616. Sosna J, Sella T, Sy O, et al. Critical analysis of the performance of double-contrast barium enema for detecting colorectal polyps > or = 6 mm in the era of CT colonography. AJR Am J Roentgenol. 2008;190(2):374-85.
617. Sotiriadis A, Makrydimas G, Ioannidis JP. Diagnostic performance of intracardiac echogenic foci for Down syndrome: a meta-analysis. Obstet Gynecol. 2003;101(5 Pt 1):1009-16.
618. Spencer-Green G, Alter D, Welch HG. Test performance in systemic sclerosis: anti-centromere and anti-Scl-70 antibodies. Am J Med. 1997;103(3):242-8.
619. Spitzmiller RE, Phillips T, Meinzen-Derr J, et al. Amplitude-integrated EEG is useful in predicting neurodevelopmental outcome in full-term infants with hypoxic-ischemic encephalopathy: a meta-analysis. J Child Neurol. 2007;22(9):1069-78.
620. St John A, Boyd JC, Lowes AJ, et al. The use of urinary dipstick tests to exclude urinary tract infection: a systematic review of the literature. Am J Clin Pathol. 2006;126(3):428-36.
621. Stein PD, Beemath A, Kayali F, et al. Multidetector computed tomography for the diagnosis of coronary artery disease: a systematic review. Am J Med. 2006;119(3):203-16.
622. Stein PD, Hull RD, Patel KC, et al. D-dimer for the exclusion of acute venous thrombosis and pulmonary embolism: a systematic review. Ann Intern Med. 2004;140(8):589-602.
623. Stein PD, Yaekoub AY, Matta F, et al. 64-slice CT for diagnosis of coronary artery disease: a systematic review. Am J Med. 2008;121(8):715-25.
624. Steingart KR, Dendukuri N, Henry M, et al. Performance of purified antigens for serodiagnosis of pulmonary tuberculosis: a meta-analysis. Clin Vaccine Immunol. 2009;16(2):260-76.
625. Steingart KR, Henry M, Laal S, et al. A systematic review of commercial serological antibody detection tests for the diagnosis of extrapulmonary tuberculosis. Postgrad Med J. 2007;83(985):705-12.

626. Steingart KR, Henry M, Laal S, et al. Commercial serological antibody detection tests for the diagnosis of pulmonary tuberculosis: a systematic review. PLoS Med. 2007;4(6):e202.
627. Steingart KR, Henry M, Ng V, et al. Fluorescence versus conventional sputum smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(9):570-81.
628. Steingart KR, Ng V, Henry M, et al. Sputum processing methods to improve the sensitivity of smear microscopy for tuberculosis: a systematic review. Lancet Infect Dis. 2006;6(10):664-74.
629. Stengel D, Bauwens K, Sehouli J, et al. Systematic review and meta-analysis of emergency ultrasonography for blunt abdominal trauma. Br J Surg. 2001;88(7):901-12.
630. Stevens C, Lee JK, Sadatsafavi M, et al**.** Pediatric thyroid fine-needle aspiration cytology: a meta-analysis. J Pediatr Surg. 2009;44(11):2184-91.
631. Sulik SM, Kroeger K, Schultz JK, et al**.** Are fluid-based cytologies superior to the conventional Papanicolaou test? A systematic review. J Fam Pract. 2001;50(12):1040-6.
632. Sun Z, Davidson R, Lin CH**.** Multi-detector row CT angiography in the assessment of coronary in-stent restenosis: a systematic review. Eur J Radiol. 2009;69(3):489-95.
633. Sun Z, Jiang W**.** Diagnostic value of multislice computed tomography angiography in coronary artery disease: a meta-analysis. Eur J Radiol. 2006;60(2):279-86.
634. Sun Z, Lin C, Davidson R, et al**.** Diagnostic value of 64-slice CT angiography in coronary artery disease: a systematic review. Eur J Radiol. 2008;67(1):78-84.
635. Swart P, Mol BW, van der Veen F, et al**.** The accuracy of hysterosalpingography in the diagnosis of tubal pathology: a meta-analysis. Fertil Steril. 1995;64(3):486-91.
636. Szadek KM, van der Wurff P, van Tulder MW, et al**.** Diagnostic validity of criteria for sacroiliac joint pain: a systematic review. J Pain. 2009;10(4):354-68.
637. Tabas JA, Rodriguez RM, Seligman HK, et al**.** Electrocardiographic criteria for detecting acute myocardial infarction in patients with left bundle branch block: a meta-analysis. Ann Emerg Med. 2008;52(4):329-36 e1.
638. Talwalkar JA, Kurtz DM, Schoenleber SJ, et a;**.** Ultrasound-based transient elastography for the detection of hepatic fibrosis: systematic review and meta-analysis. Clin Gastroenterol Hepatol. 2007;5(10):1214-20.
639. Tamariz LJ, Eng J, Segal JB, et al. Usefulness of clinical prediction rules for the diagnosis of venous thromboembolism: a systematic review. Am J Med. 2004;117(9):676-84.
640. Tan E, Anstee A, Koh DM, et al**.** Diagnostic precision of endoanal MRI in the detection of anal sphincter pathology: a meta-analysis. Int J Colorectal Dis. 2008;23(6):641-51.
641. Tan E, Gouvas N, Nicholls RJ, et al**.** Diagnostic precision of carcinoembryonic antigen in the detection of recurrence of colorectal cancer. Surg Oncol. 2009;18(1):15-24.
642. Tan KT, van Beek EJ, Brown PW, et al**.** Magnetic resonance angiography for the diagnosis of renal artery stenosis: a meta-analysis. Clin Radiol. 2002;57(7):617-24.
643. Tandon S, Shahab R, Benton JI, et al**.** Fine-needle aspiration cytology in a regional head and neck cancer center: comparison with a systematic review and meta-analysis. Head Neck. 2008;30(9):1246-52.
644. Tang BM, Eslick GD, Craig JC, et al**.** Accuracy of procalcitonin for sepsis diagnosis in critically ill patients: systematic review and meta-analysis. Lancet Infect Dis. 2007;7(3):210-7.
645. Tebas P, Nease RF, Storch GA**.** Use of the polymerase chain reaction in the diagnosis of herpes simplex encephalitis: a decision analysis model. Am J Med. 1998;105(4):287-95.
646. Temmerman OP, Heyligers IC, Teule GJ, et al**.** The value of contrast and subtraction arthrography in the assessment of aseptic loosening of total hip prostheses: a meta-analysis. Eur J Radiol. 2005;56(1):113-9.

647. Temmerman OP, Raijmakers PG, Berkhof J, et al. Accuracy of diagnostic imaging techniques in the diagnosis of aseptic loosening of the femoral component of a hip prosthesis: a meta-analysis. J Bone Joint Surg Br. 2005;87(6):781-5.
648. Temmerman OP, Raijmakers PG, Deville WL, et. The use of plain radiography, subtraction arthrography, nuclear arthrography, and bone scintigraphy in the diagnosis of a loose acetabular component of a total hip prosthesis: a systematic review. J Arthroplasty. 2007;22(6):818-27.
649. Teng CL, Ng CJ, Nik-Sherina H, et al. The accuracy of mother's touch to detect fever in children: a systematic review. J Trop Pediatr. 2008;54(1):70-3.
650. Tenner S, Dubner H, Steinberg W. Predicting gallstone pancreatitis with laboratory parameters: a meta-analysis. Am J Gastroenterol. 1994;89(10):1863-6.
651. Terasawa T, Blackmore CC, Bent S, et al. Systematic review: computed tomography and ultrasonography to detect acute appendicitis in adults and adolescents. Ann Intern Med. 2004;141(7):537-46.
652. Terasawa T, Lau J, Bardet S, et al. Fluorine-18-fluorodeoxyglucose positron emission tomography for interim response assessment of advanced-stage Hodgkin's lymphoma and diffuse large B-cell lymphoma: a systematic review. J Clin Oncol. 2009;27(11):1906-14.
653. Terasawa T, Nihashi T, Hotta T, et al. 18F-FDG PET for posttherapy assessment of Hodgkin's disease and aggressive non-Hodgkin's lymphoma: a systematic review. J Nucl Med. 2008;49(1):13-21.
654. Termaat MF, Raijmakers PG, Scholten HJ, et al. The accuracy of diagnostic imaging for the assessment of chronic osteomyelitis: a systematic review and meta-analysis. J Bone Joint Surg Am. 2005;87(11):2464-71.
655. Tew K, Irwig L, Matthews A, et al. Meta-analysis of sentinel node imprint cytology in breast cancer. Br J Surg. 2005;92(9):1068-80.
656. Thangaratinam S, Coomarasamy A, O'Mahony F, Sharp S, Zamora J, Khan KS, et al. Estimation of proteinuria as a predictor of complications of pre-eclampsia: a systematic review. BMC Med. 2009;7:10.
657. Thangaratinam S, Daniels J, Ewer AK, et al. Accuracy of pulse oximetry in screening for congenital heart disease in asymptomatic newborns: a systematic review. Arch Dis Child Fetal Neonatal Ed. 2007;92(3):F176-80.
658. Thangaratinam S, Ismail KM, Sharp S, et al. Accuracy of serum uric acid in predicting complications of pre-eclampsia: a systematic review. BJOG. 2006;113(4):369-78.
659. Thomas SM, Goodacre SW, Sampson FC, et al. Diagnostic value of CT for deep vein thrombosis: results of a systematic review and meta-analysis. Clin Radiol. 2008;63(3):299-304.
660. Thorell LH. Valid electrodermal hyporeactivity for depressive suicidal propensity offers links to cognitive theory. Acta Psychiatr Scand. 2009;119(5):338-49.
661. Trochez-Martinez RD, Smith P, Lamont RF. Use of C-reactive protein as a predictor of chorioamnionitis in preterm prelabour rupture of membranes: a systematic review. BJOG. 2007;114(7):796-801.
662. Trowbridge RL, Rutkowski NK, Shojania KG. Does this patient have acute cholecystitis? JAMA. 2003;289(1):80-6.
663. Tsao H, Nadiminti U, Sober AJ, et al. A meta-analysis of reverse transcriptase-polymerase chain reaction for tyrosinase mRNA as a marker for circulating tumor cells in cutaneous melanoma. Arch Dermatol. 2001;137(3):325-30.
664. Tse F, Liu L, Barkun AN, et al. EUS: a meta-analysis of test performance in suspected choledocholithiasis. Gastrointest Endosc. 2008;67(2):235-44.
665. Tsushima Y, Takahashi-Taketomi A, Endo K. Magnetic resonance (MR) differential diagnosis of breast tumors using apparent diffusion coefficient (ADC) on 1.5-T. J Magn Reson Imaging. 2009;30(2):249-55.

666. Tuon FF. A systematic literature review on the diagnosis of invasive aspergillosis using polymerase chain reaction (PCR) from bronchoalveolar lavage clinical samples. Rev Iberoam Micol. 2007;24(2):89-94.
667. Tuon FF, Litvoc MN, Lopes MI. Adenosine deaminase and tuberculous pericarditis—a systematic review with meta-analysis. Acta Trop. 2006;99(1):67-74.
668. Urbach DR, Khajanchee YS, Jobe BA, et al. Cost-effective management of common bile duct stones: a decision analysis of the use of endoscopic retrograde cholangiopancreatography (ERCP), intraoperative cholangiography, and laparoscopic bile duct exploration. Surg Endosc. 2001;15(1):4-13.
669. Uzzan B, Cohen R, Nicolas P, et al. Procalcitonin as a diagnostic test for sepsis in critically ill adults and after surgery or trauma: a systematic review and meta-analysis. Crit Care Med. 2006;34(7):1996-2003.
670. Vakil N, Moayyedi P, Fennerty MB, et al. Limited value of alarm features in the diagnosis of upper gastrointestinal malignancy: systematic review and meta-analysis. Gastroenterology. 2006;131(2):390-401; quiz 659-60.
671. van Beek EJ, Brouwers EM, Song B, et al. Lung scintigraphy and helical computed tomography for the diagnosis of pulmonary embolism: a meta-analysis. Clin Appl Thromb Hemost. 2001;7(2):87-92.
672. van de Laar R, van der Ham DP, Oei SG, et al. Accuracy of C-reactive protein determination in predicting chorioamnionitis and neonatal infection in pregnant women with premature rupture of membranes: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2009;147(2):124-9.
673. van de Lande J, Torrenga B, Raijmakers PG, et al. Sentinel lymph node detection in early stage uterine cervix carcinoma: a systematic review. Gynecol Oncol. 2007;106(3):604-13.
674. van den Broek FJ, Reitsma JB, Curvers WL et al. Systematic review of narrow-band imaging for the detection and differentiation of neoplastic and nonneoplastic lesions in the colon (with videos). Gastrointest Endosc. 2009;69(1):124-35.
675. Van Der Horst-Schrivers AN, Jager PL, Boezen HM, et al. Iodine-123 metaiodobenzylguanidine scintigraphy in localising phaeochromocytomas—experience and meta-analysis. Anticancer Res. 2006;26(2B):1599-604.
676. van der Meer V, Neven AK, van den Broek PJ, et al. Diagnostic value of C reactive protein in infections of the lower respiratory tract: systematic review. BMJ. 2005;331(7507):26.
677. van der Ploeg IM, Nieweg OE, van Rijk MC, et al. Axillary recurrence after a tumour-negative sentinel node biopsy in breast cancer patients: A systematic review and meta-analysis of the literature. Eur J Surg Oncol. 2008;34(12):1277-84.
678. van der Zaag-Loonen HJ, Dikkers R, de Bock GH, et al. The clinical value of a negative multi-detector computed tomographic angiography in patients suspected of coronary artery disease: A meta-analysis. Eur Radiol. 2006;16(12):2748-56.
679. van Deurzen CH, Vriens BE, Tjan-Heijnen VC, et al. Accuracy of sentinel node biopsy after neoadjuvant chemotherapy in breast cancer patients: a systematic review. Eur J Cancer. 2009;45(18):3124-30.
680. van Dongen H, de Kroon CD, Jacobi CE, et al. Diagnostic hysteroscopy in abnormal uterine bleeding: a systematic review and meta-analysis. BJOG. 2007;114(6):664-75.
681. van Gelder JM. Computed tomographic angiography for detecting cerebral aneurysms: implications of aneurysm size distribution for the sensitivity, specificity, and likelihood ratios. Neurosurgery. 2003;53(3):597-605; discussion -6.
682. van Randen A, Bipat S, Zwinderman AH, et al. Acute appendicitis: meta-analysis of diagnostic performance of CT and graded compression US related to prevalence of disease. Radiology. 2008;249(1):97-106.

683. van Rijkom HM, Verdonschot EH. Factors involved in validity measurements of diagnostic tests for approximal caries—a meta-analysis. Caries Res. 1995;29(5):364-70.
684. van Schaik IN, Bossuyt PM, Brand A, et al. Diagnostic value of GM1 antibodies in motor neuron disorders and neuropathies: a meta-analysis. Neurology. 1995;45(8):1570-7.
685. van Vliet EP, Heijenbrok-Kal MH, Hunink MG, et al. Staging investigations for oesophageal cancer: a meta-analysis. Br J Cancer. 2008;98(3):547-57.
686. van Weert JM, Repping S, Van Voorhis BJ, et al. Performance of the postwash total motile sperm count as a predictor of pregnancy at the time of intrauterine insemination: a meta-analysis. Fertil Steril. 2004;82(3):612-20.
687. van Westreenen HL, Westerterp M, Bossuyt PM, et al. Systematic review of the staging performance of 18F-fluorodeoxyglucose positron emission tomography in esophageal cancer. J Clin Oncol. 2004;22(18):3805-12.
688. Van Zaane B, Zuithoff NP, Reitsma JB, et al. Meta-analysis of the diagnostic accuracy of transesophageal echocardiography for assessment of atherosclerosis in the ascending aorta in patients undergoing cardiac surgery. Acta Anaesthesiol Scand. 2008;52(9):1179-87.
689. Vanezis AP, Bhopal R. Validity of electrocardiographic classification of left ventricular hypertrophy across adult ethnic groups with echocardiography as a standard. J Electrocardiol. 2008;41(5):404-12.
690. Vanhoenacker PK, Decramer I, Bladt O, et al. Detection of non-ST-elevation myocardial infarction and unstable angina in the acute setting: meta-analysis of diagnostic performance of multi-detector computed tomographic angiography. BMC Cardiovasc Disord. 2007;7:39.
691. Vanhoenacker PK, Decramer I, Bladt O, et al. Multidetector computed tomography angiography for assessment of in-stent restenosis: meta-analysis of diagnostic performance. BMC Med Imaging. 2008;8:14.
692. Vanhoenacker PK, Heijenbrok-Kal MH, Van Heste R, et al. Diagnostic performance of multidetector CT angiography for assessment of coronary artery disease: meta-analysis. Radiology. 2007;244(2):419-28.
693. Varonen H, Makela M, Savolainen S, et al. Comparison of ultrasound, radiography, and clinical examination in the diagnosis of acute maxillary sinusitis: a systematic review. J Clin Epidemiol. 2000;53(9):940-8.
694. Vasbinder GB, Nelemans PJ, Kessels AG, et al. Diagnostic tests for renal artery stenosis in patients suspected of having renovascular hypertension: a meta-analysis. Ann Intern Med. 2001;135(6):401-11.
695. Vasquez TE, Rimkus DS, Hass MG, et al. Efficacy of morphine sulfate-augmented hepatobiliary imaging in acute cholecystitis. J Nucl Med Technol. 2000;28(3):153-5.
696. Vercammen MJ, Verloes A, Van de Velde H, et al. Accuracy of soluble human leukocyte antigen-G for predicting pregnancy among women undergoing infertility treatment: meta-analysis. Hum Reprod Update. 2008;14(3):209-18.
697. Verkooijen HM, Peeters PH, Buskens E, et al. Diagnostic accuracy of large-core needle biopsy for nonpalpable breast disease: a meta-analysis. Br J Cancer. 2000;82(5):1017-21.
698. Verma D, Kapadia A, Eisen GM, et al. EUS vs MRCP for detection of choledocholithiasis. Gastrointest Endosc. 2006;64(2):248-54.
699. Vestergaard ME, Macaskill P, Holt PE, et al. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br J Dermatol. 2008;159(3):669-76.
700. Vijayasankar D, Boyle AA, Atkinson P. Can the Ottawa knee rule be applied to children? A systematic review and meta-analysis of observational studies. Emerg Med J. 2009;26(4):250-3.

701. Virgili G, Menchini F, Dimastrogiovanni AF, et al. Optical coherence tomography versus stereoscopic fundus photography or biomicroscopy for diagnosing diabetic macular edema: a systematic review. Invest Ophthalmol Vis Sci. 2007;48(11):4963-73.
702. Visser K, Hunink MG. Peripheral arterial disease: gadolinium-enhanced MR angiography versus color-guided duplex US—a meta-analysis. Radiology. 2000;216(1):67-77.
703. Vlaar AM, Bouwmans A, Mess WH, et al. Transcranial duplex in the differential diagnosis of parkinsonian syndromes: a systematic review. J Neurol. 2009;256(4):530-8.
704. Vlaar AM, van Kroonenburgh MJ, Kessels AG, et al. Meta-analysis of the literature on diagnostic accuracy of SPECT in parkinsonian syndromes. BMC Neurol. 2007;7:27.
705. von Roon AC, Karamountzos L, Purkayastha S, et al. Diagnostic precision of fecal calprotectin for inflammatory bowel disease and colorectal malignancy. Am J Gastroenterol. 2007;102(4):803-13.
706. Vroomen PC, de Krom MC, Knottnerus JA. Diagnostic value of history and physical examination in patients suspected of sciatica due to disc herniation: a systematic review. J Neurol. 1999;246(10):899-906.
707. Wald DS, Bestwick JP. Carotid ultrasound screening for coronary heart disease: results based on a meta-analysis of 18 studies and 44,861 subjects. J Med Screen. 2009;16(3):147-54.
708. Wald DS, Bestwick JP, Wald NJ. Child-parent screening for familial hypercholesterolaemia: screening strategy based on a meta-analysis. BMJ. 2007;335(7620):599.
709. Wald NJ, Kennard A, Hackshaw AK. First trimester serum screening for Down's syndrome. Prenat Diagn. 1995;15(13):1227-40.
710. Walleser S, Griffiths A, Lord SJ, et al. What is the value of computered tomography colonography in patients screening positive for fecal occult blood? A systematic review and economic evaluation. Clin Gastroenterol Hepatol. 2007;5(12):1439-46; quiz 368.
711. Walton DM, Sadi J. Identifying SLAP lesions: a meta-analysis of clinical tests and exercise in clinical reasoning. Phys Ther Sport. 2008;9(4):167-76.
712. Wan Y, Xu YY, Jiang JH, et al. Chinese literature associated with diagnosis of Helicobacter pylori. World J Gastroenterol. 2004;10(2):231-3.
713. Wancata J, Alexandrowicz R, Marquart B, et al. The criterion validity of the Geriatric Depression Scale: a systematic review. Acta Psychiatr Scand. 2006;114(6):398-410.
714. Wang P, Guo YM, Liu M, et al. A meta-analysis of the accuracy of prostate cancer studies which use magnetic resonance spectroscopy as a diagnostic tool. Korean J Radiol. 2008;9(5):432-8.
715. Wang WH, Huang JQ, Zheng GF, et al. Is proton pump inhibitor testing an effective approach to diagnose gastroesophageal reflux disease in patients with noncardiac chest pain?: a meta-analysis. Arch Intern Med. 2005;165(11):1222-8.
716. Wang Y, Sun G, Pan JG, et al. Performance of tPSA and f/tPSA for prostate cancer in Chinese. A systematic review and meta-analysis. Prostate Cancer Prostatic Dis. 2006;9(4):374-8.
717. Wang ZP, Li H, Hao LZ, et al. The effectiveness of prenatal serum biomarker screening for neural tube defects in second trimester pregnant women: a meta-analysis. Prenat Diagn. 2009;29(10):960-5.
718. Wardlaw JM, Chappell FM, Best JJ, et al. Non-invasive imaging compared with intra-arterial angiography in the diagnosis of symptomatic carotid stenosis: a meta-analysis. Lancet. 2006;367(9521):1503-12.
719. Warner E, Messersmith H, Causer P, et al. Systematic review: using magnetic resonance imaging to screen women at high risk for breast cancer. Ann Intern Med. 2008;148(9):671-9.

720. Watson EJ, Templeton A, Russell I, et al. The accuracy and efficacy of screening tests for Chlamydia trachomatis: a systematic review. J Med Microbiol. 2002;51(12):1021-31.
721. Weinschenker P, Soares HP, Clark O, et al**.** Immunocytochemical detection of epithelial cells in the bone marrow of primary breast cancer patients: a meta-analysis. Breast Cancer Res Treat. 2004;87(3):215-24.
722. Wellnitz U, Binder B, Fritz P, et al**.** Reliability of telepathology for frozen section service. Anal Cell Pathol. 2000;21(3-4):213-22.
723. Wells PS, Lensing AW, Davidson BL, et al**.** Accuracy of ultrasound for the diagnosis of deep venous thrombosis in asymptomatic patients after orthopedic surgery. A meta-analysis. Ann Intern Med. 1995;122(1):47-53.
724. West J, Goodacre S, Sampson F**.** The value of clinical features in the diagnosis of acute pulmonary embolism: systematic review and meta-analysis. QJM. 2007;100(12):763-9.
725. Westerterp M, van Westreenen HL, Reitsma JB, et al. Esophageal cancer: CT, endoscopic US, and FDG PET for assessment of response to neoadjuvant therapy—systematic review. Radiology. 2005;236(3):841-51.
726. Weston AR, Jackson TJ, Blamey S**.** Diagnosis of appendicitis in adults by ultrasonography or computed tomography: a systematic review and meta-analysis. Int J Technol Assess Health Care. 2005;21(3):368-79.
727. Westwood ME, Kelly S, Berry E, et al. Use of magnetic resonance angiography to select candidates with recently symptomatic carotid stenosis for surgery: systematic review. BMJ. 2002;324(7331):198.
728. Westwood ME, Whiting PF, Cooper J, et al**.** Further investigation of confirmed urinary tract infection (UTI) in children under five years: a systematic review. BMC Pediatr. 2005;5(1):2.
729. White PM, Wardlaw JM, Easton V**.** Can noninvasive imaging accurately depict intracranial aneurysms? A systematic review. Radiology. 2000;217(2):361-70.
730. White RH, McGahan JP, Daschbach MM, et al**.** Diagnosis of deep-vein thrombosis using duplex ultrasound. Ann Intern Med. 1989;111(4):297-304.
731. Whiting P, Harbord R, Main C, et al. Accuracy of magnetic resonance imaging for the diagnosis of multiple sclerosis: systematic review. BMJ. 2006;332(7546):875-84.
732. Whiting P, Westwood M, Watt I, et al**.** Rapid tests and urine sampling techniques for the diagnosis of urinary tract infection (UTI) in children under five years: a systematic review. BMC Pediatr. 2005;5(1):4.
733. Whitsel EA, Boyko EJ, Siscovick DS**.** Reassessing the role of QTc in the diagnosis of autonomic failure among patients with diabetes: a meta-analysis. Diabetes Care. 2000;23(2):241-7.
734. Wiering B, Krabbe PF, Jager GJ, et al**.** The impact of fluor-18-deoxyglucose-positron emission tomography in the management of colorectal liver metastases. Cancer. 2005;104(12):2658-70.
735. Wiese W, Patel SR, Patel SC, et al**.** A meta-analysis of the Papanicolaou smear and wet mount for the diagnosis of vaginal trichomoniasis. Am J Med. 2000;108(4):301-8.
736. Wijnberger LD, Huisjes AJ, Voorbij HA, et al**.** The accuracy of lamellar body count and lecithin/sphingomyelin ratio in the prediction of neonatal respiratory distress syndrome: a meta-analysis. BJOG. 2001;108(6):583-8.
737. Wilhelmus KR, Hassan SS**.** The prognostic role of donor corneoscleral rim cultures in corneal transplantation. Ophthalmology. 2007;114(3):440-5.
738. Will O, Purkayastha S, Chan C, et al. Diagnostic precision of nanoparticle-enhanced MRI for lymph-node metastases: a meta-analysis. Lancet Oncol. 2006;7(1):52-60.
739. Williams GJ, Macaskill P, Chan SF, et al. Comparative accuracy of renal duplex sonographic parameters in the diagnosis of renal artery stenosis: paired and unpaired analysis. AJR Am J Roentgenol. 2007;188(3):798-811.

740. Williams JW, Jr., Noel PH, Cordes JA, et al**.** Is this patient clinically depressed? JAMA. 2002;287(9):1160-70.
741. Willmann O, Wennberg R, May T, Woermann FG, Pohlmann-Eden B**.** The role of 1H magnetic resonance spectroscopy in pre-operative evaluation for epilepsy surgery. A meta-analysis. Epilepsy Res. 2006;71(2-3):149-58.
742. Willmann O, Wennberg R, May T, et al**.** The contribution of 18F-FDG PET in preoperative epilepsy surgery evaluation for patients with temporal lobe epilepsy A meta-analysis. Seizure. 2007;16(6):509-20.
743. Windeler J, Kobberling J**.** Colorectal carcinoma and Haemoccult. A study of its value in mass screening using meta-analysis. Int J Colorectal Dis. 1987;2(4):223-8.
744. Wittkampf KA, Naeije L, Schene AH, et al**.** Diagnostic accuracy of the mood module of the Patient Health Questionnaire: a systematic review. Gen Hosp Psychiatry. 2007;29(5):388-95.
745. Wiwanitkit V**.** Prostate specific antigen for screening for prostate cancer: an appraisal of Thai reports. Asian Pac J Cancer Prev. 2004;5(4):406-8.
746. Wolfer LR, Derby R, Lee JE, et al**.** Systematic review of lumbar provocation discography in asymptomatic subjects with a meta-analysis of false-positive rates. Pain Physician. 2008;11(4):513-38.
747. Worster A, Balion CM, Hill SA, et al. Diagnostic accuracy of BNP and NT-proBNP in patients presenting to acute care settings with dyspnea: a systematic review. Clin Biochem. 2008;41(4-5):250-9.
748. Worster A, Preyra I, Weaver B, et al**.** The accuracy of noncontrast helical computed tomography versus intravenous pyelography in the diagnosis of suspected acute urolithiasis: a meta-analysis. Ann Emerg Med. 2002;40(3):280-6.
749. Wu AH, Lane PL**.** Metaanalysis in clinical chemistry: validation of cardiac troponin T as a marker for ischemic heart diseases. Clin Chem. 1995;41(8 Pt 2):1228-33.
750. Wykes CB, Clark TJ, Khan KS**.** Accuracy of laparoscopy in the diagnosis of endometriosis: a systematic quantitative review. BJOG. 2004;111(11):1204-12.
751. Xing Y, Foy M, Cox DD, et al**.** Meta-analysis of sentinel lymph node biopsy after preoperative chemotherapy in patients with breast cancer. Br J Surg. 2006;93(5):539-46.
752. Yu SH, Kim CB, Park JW, et al**.** Ultrasonography in the diagnosis of appendicitis: evaluation by meta-analysis. Korean J Radiol. 2005;6(4):267-77.
753. Yuan W, Chen L, Bernal AL**.** Is elevated maternal serum alpha-fetoprotein in the second trimester of pregnancy associated with increased preterm birth risk? A systematic review and meta-analysis. Eur J Obstet Gynecol Reprod Biol. 2009;145(1):57-64.
754. Yuan Y, Gu ZX, Wei WS**.** Fluorodeoxyglucose-positron-emission tomography, single-photon emission tomography, and structural MR imaging for prediction of rapid conversion to Alzheimer disease in patients with mild cognitive impairment: a meta-analysis. AJNR Am J Neuroradiol. 2009;30(2):404-10.
755. Zandbergen EG, de Haan RJ, Stoutenbeek CP, et al**.** Systematic review of early prediction of poor outcome in anoxic-ischaemic coma. Lancet. 1998;352(9143):1808-12.
756. Zhang C, Chen Y, Xue H, et al. Diagnostic value of FDG-PET in recurrent colorectal carcinoma: a meta-analysis. Int J Cancer. 2009;124(1):167-73.
757. Zhao WY, Luo M, Sun YW, et al. Computed tomography in diagnosing vascular invasion in pancreatic and periampullary cancers: a systematic review and meta-analysis. Hepatobiliary Pancreat Dis Int. 2009;8(5):457-64.
758. Zielinski GD, Bais AG, Helmerhorst TJ, et al. HPV testing and monitoring of women after treatment of CIN 3: review of the literature and meta-analysis. Obstet Gynecol Surv. 2004;59(7):543-53.

759. Zijlstra JM, Lindauer-van der Werf G, Hoekstra OS, et al**.** 18F-fluoro-deoxyglucose positron emission tomography for post-treatment evaluation of malignant lymphoma: a systematic review. Haematologica. 2006;91(4):522-9.

760. Zintzaras E, Germenis AE**.** Performance of antibodies against tissue transglutaminase for the diagnosis of celiac disease: meta-analysis. Clin Vaccine Immunol. 2006;13(2):187-92.

Appendix D. Reasons for Exclusion

Reasons for Exclusion	**Studies**
Not meta-analyses of test accuracy	319
Cochrane Reviews/ Health Technology Assessments	32
Individual Patient Data meta-analyses	14
Narrative reviews	54
Editorials/ letters/ commentaries	17
Methodological contributions	13
Risk score meta-analyses	4
Meta-analyses of continuous outcomes	4
No gold standard	4
Duplicate publications	3
Animal study	1
TOTAL	***465***

Appendix E. Regression Analyses for Trends Over Time

Appendix Table E1. Regression analyses for trends over time, for all studies included in the overview (publication years 1987 to 2009)

Characteristic	Per Year OR	P-value
Cardiovascular disease	0.99 (0.95, 1.03)	0.585
Obstetrics and gynecology	0.98 (0.94, 1.02)	0.352
Gastrointestinal disease	1.07 (1.02, 1.13)	0.008
Infectious disease	1.01 (0.97, 1.06)	0.581
Oncology	1.05 (1.01, 1.09)	0.02
Nephrology and urology	0.99 (0.92, 1.07)	0.86
Rheumatology	1.05 (0.88, 1.26)	0.559
Pulmonary medicine	1.10 (1.03, 1.18)	0.005
Orthopedics	1.09 (1.00, 1.19)	0.041
Psychiatry	1.12 (0.99, 1.27)	0.077
Ear-nose-throat	1.02 (0.92, 1.13)	0.732
Neurology	1.02 (0.95, 1.10)	0.591
Pediatrics	1.11 (1.00, 1.24)	0.057
Histology	0.99 (0.94, 1.03)	0.58
Clinical exam	1.00 (0.96, 1.05)	0.857
Imaging	1.00 (0.97, 1.03)	0.905
Biomarker	1.02 (0.98, 1.06)	0.324
Clinical challenge tests	1.11 (1.00, 1.23)	0.046
Physiologic tests	0.99 (0.92, 1.06)	0.667
Endoscopy	1.03 (0.93, 1.15)	0.54
Exact search described	1.09 (1.05, 1.14)	<0.001
Search terms provided	1.01 (0.98, 1.05)	0.424
Search on demand (on non-journal website or from the authors)	1.03 (0.96, 1.11)	0.367
Years searched were reported	1.08 (1.02, 1.13)	0.004
Quality criteria for study selection	1.06 (1.02, 1.11)	0.007
Consideration of English-language studies only	0.96 (0.93, 1.00)	0.028
Consideration of at least one language other than English	1.05 (1.02, 1.09)	0.003
Medline	1.19 (1.11, 1.27)	<0.001
EMBASE	1.28 (1.23, 1.35)	<0.001
Conference proceedings	1.08 (1.02, 1.15)	0.007
Reference lists of eligible studies	1.02 (0.98, 1.06)	0.315
Reference lists of relevant review articles	1.01 (0.97, 1.04)	0.672
SCI or other ISI databases	1.18 (1.09, 1.28)	<0.001
Current Contents	0.92 (0.86, 0.98)	0.015
Experts in the field	0.98 (0.94, 1.02)	0.311
Test manufacturers	1.05 (0.96, 1.15)	0.259
Other specific electronic databases	1.25 (1.18, 1.32)	<0.001
Hand searching of journals	0.99 (0.95, 1.04)	0.664
Cochrane databases	1.30 (1.23, 1.38)	<0.001
CINAHL	1.25 (1.15, 1.36)	<0.001
At least one electronic database in addition to Medline	1.30 (1.24, 1.35)	<0.001

Unpublished information	0.97 (0.93, 1.02)	0.212
Any quality assessment	1.34 (1.28, 1.40)	<0.001
Study settings	1.03 (0.99, 1.06)	0.114
Consecutive patient recruitment	1.07 (1.04, 1.11)	<0.001
Prospective study design	1.12 (1.09, 1.16)	<0.001
Details of the reference standard test	0.98 (0.93, 1.03)	0.487
Reader expertise	1.12 (1.06, 1.19)	<0.001
Blinding (index test assessor to reference standard results)	1.09 (0.98, 1.21)	0.102
Blinding (reference standard test assessor to index test results)	1.09 (0.98, 1.22)	0.094
Blinding (unspecified)	0.87 (0.75, 1.00)	0.057
Any blinding	1.09 (1.05, 1.13)	<0.001
Patient age	1.15 (1.11, 1.20)	<0.001
Patient sex	1.06 (0.95, 1.18)	0.274
Location of primary studies	1.20 (1.14, 1.27)	<0.001
Spectrum bias	1.16 (1.04, 1.29)	0.007
Selection bias	1.14 (1.03, 1.27)	0.015
Time between index and reference standard test application	1.31 (1.17, 1.46)	<0.001
Test independence	1.19 (1.07, 1.33)	0.002
Indeterminate test results	1.40 (1.24, 1.58)	<0.001
Withdrawals	1.51 (1.33, 1.71)	<0.001
Verification bias	1.17 (1.13, 1.22)	<0.001
QUADAS	1.76 (1.57, 1.99)	<0.001
STARD	1.32 (1.19, 1.47)	<0.001
Random effects used	1.21 (1.16, 1.26)	<0.001
Bayesian analyses	1.08 (0.95, 1.22)	0.261
OR	1.17 (1.12, 1.22)	<0.001
Sensitivity	1.07 (1.03, 1.11)	<0.001
Specificity	1.08 (1.04, 1.12)	<0.001
Likelihood ratios	1.13 (1.08, 1.18)	<0.001
Accuracy	0.97 (0.91, 1.04)	0.359
Predictive values	0.99 (0.95, 1.04)	0.735
Q*	0.91 (0.84, 0.97)	0.007
AUC	0.99 (0.92, 1.06)	0.721
Any graphical display of analysis (synthesis) results	1.13 (1.09, 1.17)	<0.001
Forest plots	1.28 (1.22, 1.35)	<0.001
ROC space plots	1.06 (1.03, 1.10)	<0.001
Heterogeneity testing	1.21 (1.17, 1.26)	<0.001
Univariate analyses	0.82 (0.76, 0.88)	<0.001
Comparative analyses	0.98 (0.94, 1.02)	0.292
Advanced synthesis methods	1.27 (1.16, 1.39)	<0.001
Any exploration of heterogeneity	1.08 (1.04, 1.12)	<0.001
Provides data for re-analyses	0.96 (0.93, 1.00)	0.027

Results are limited to 2004 onwards for the following variables: blinding (index test assessor to reference standard results, or reference standard assessor to index test results, or unspecified); selection bias; spectrum bias; extraction of data on participants’ sex; withdrawals; indeterminate test results; timing; test independence. The STARD and QUADAS checklists were first published in January 2003 and November 2003, respectively.

AUC = area under the curve; CINAHL = Cumulative Index to Nursing and Allied Health Literature; OR = odds ratio; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; ROC = receiver operating characteristic; STARD = Standards for Reporting of Diagnostic Accuracy.

Appendix Table E2. Regression analyses for trends over time, for studies included in the overview published since 2005 (publication years 2005 to 2009)

Characteristic	Per Year OR	P-value
Cardiovascular disease	0.98 (0.83, 1.15)	0.767
Obstetrics and gynecology	1.18 (0.96, 1.44)	0.117
Gastrointestinal disease	0.92 (0.78, 1.09)	0.331
Infectious disease	0.87 (0.72, 1.05)	0.141
Oncology	1.05 (0.90, 1.23)	0.498
Nephrology and urology	0.78 (0.58, 1.06)	0.113
Rheumatology	1.25 (0.64, 2.46)	0.514
Pulmonary medicine	1.07 (0.86, 1.33)	0.543
Orthopedics	1.08 (0.82, 1.40)	0.589
Psychiatry	1.16 (0.79, 1.71)	0.439
Ear-nose-throat	0.87 (0.58, 1.30)	0.488
Neurology	1.15 (0.83, 1.61)	0.398
Pediatrics	1.01 (0.75, 1.37)	0.929
Histology	1.01 (0.83, 1.24)	0.899
Clinical exam	1.00 (0.83, 1.22)	0.978
Imaging	0.97 (0.85, 1.11)	0.647
Biomarker	1.01 (0.87, 1.17)	0.905
Clinical challenge tests	0.95 (0.69, 1.30)	0.743
Physiologic tests	0.81 (0.60, 1.11)	0.200
Endoscopy	1.31 (0.84, 2.05)	0.237
Exact search described	1.18 (1.02, 1.37)	0.027
Search terms provided	0.94 (0.82, 1.08)	0.410
Search on demand (on non-journal website or from the authors)	0.99 (0.73, 1.34)	0.948
Years searched were reported	1.13 (0.86, 1.49)	0.372
Quality criteria for study selection	1.13 (0.96, 1.34)	0.151
Consideration of English-language studies only	1.06 (0.92, 1.22)	0.450
Consideration of at least one language other than English	1.01 (0.88, 1.15)	0.912
Medline	1.52 (0.95, 2.44)	0.079
EMBASE	1.02 (0.89, 1.17)	0.778
Conference proceedings	1.13 (0.92, 1.38)	0.242
Reference lists of eligible studies	0.96 (0.82, 1.13)	0.652
Reference lists of relevant review articles	0.87 (0.75, 1.01)	0.075
SCI or other ISI databases	1.14 (0.94, 1.39)	0.189
Current Contents	1.16 (0.77, 1.75)	0.480
Experts in the field	0.84 (0.70, 1.02)	0.073
Test manufacturers	0.69 (0.49, 0.96)	0.027
Other specific electronic databases	1.14 (0.99, 1.31)	0.065
Hand searching of journals	0.74 (0.60, 0.91)	0.004
Cochrane databases	0.99 (0.86, 1.13)	0.861
CINAHL	0.98 (0.81, 1.18)	0.818
At least one electronic database in addition to Medline	1.20 (1.02, 1.41)	0.031
Unpublished information	0.84 (0.68, 1.05)	0.123
Any quality assessment	1.29 (1.03, 1.63)	0.028
Study settings	1.08 (0.94, 1.25)	0.261
Consecutive patient recruitment	0.89 (0.77, 1.02)	0.084

Prospective study design	0.84 (0.73, 0.98)	0.025
Details of the reference standard test	1.09 (0.89, 1.32)	0.417
Reader expertise	1.09 (0.92, 1.30)	0.312
Blinding (index test assessor to reference standard results)	1.12 (0.97, 1.28)	0.113
Blinding (reference standard test assessor to index test results)	1.12 (0.98, 1.28)	0.106
Blinding (unspecified)	0.78 (0.65, 0.95)	0.011
Any blinding	0.92 (0.79, 1.07)	0.292
Patient age	1.09 (0.95, 1.25)	0.210
Patient sex	1.09 (0.95, 1.25)	0.224
Location of primary studies	1.07 (0.92, 1.23)	0.385
Spectrum bias	1.16 (1.02, 1.33)	0.029
Selection bias	1.11 (0.96, 1.27)	0.148
Time between index and reference standard test application	1.33 (1.16, 1.54)	<0.001
Test independence	1.10 (0.96, 1.26)	0.189
Indeterminate test results	1.29 (1.12, 1.49)	0.001
Withdrawals	1.42 (1.22, 1.65)	<0.001
Verification bias	0.97 (0.84, 1.11)	0.625
QUADAS	1.46 (1.25, 1.71)	<0.001
STARD	1.03 (0.84, 1.25)	0.789
Random effects used	1.13 (0.97, 1.31)	0.113
Bayesian analyses	1.12 (0.70, 1.78)	0.634
OR	1.21 (1.05, 1.39)	0.007
Sensitivity	1.14 (0.96, 1.35)	0.144
Specificity	1.18 (1.00, 1.39)	0.046
Likelihood ratios	1.09 (0.95, 1.26)	0.221
Accuracy	0.93 (0.67, 1.28)	0.640
Predictive values	1.05 (0.86, 1.28)	0.639
Q*	0.99 (0.55, 1.76)	0.964
AUC	1.39 (0.93, 2.06)	0.109
Any graphical display of analysis (synthesis) results	1.13 (0.94, 1.35)	0.190
Forest plots	1.26 (1.09, 1.44)	0.001
ROC space plots	0.99 (0.87, 1.14)	0.920
Heterogeneity testing	1.03 (0.88, 1.20)	0.743
Univariate analyses	0.95 (0.79, 1.13)	0.543
Comparative analyses	0.84 (0.70, 1.00)	0.056
Advanced synthesis methods	1.39 (1.13, 1.71)	0.002
Any exploration of heterogeneity	0.99 (0.86, 1.14)	0.862
Provides data for re-analyses	0.95 (0.83, 1.09)	0.458

Results are presented for 2005 onwards for all variables.
AUC = area under the curve; CINAHL = Cumulative Index to Nursing and Allied Health Literature; OR = odds ratio; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; ROC = receiver operating characteristic; STARD = Standards for Reporting of Diagnostic Accuracy.

Appendix Table E3. Regression analyses comparing meta-analyses before and after the introduction of QUADAS and STARD (2003 or later versus 2002 or earlier)

Characteristic	Per Year OR	P-value
Cardiovascular disease	0.95 (0.65, 1.40)	0.805
Obstetrics and gynecology	0.84 (0.54, 1.28)	0.409
Gastrointestinal disease	2.62 (1.54, 4.44)	<0.001
Infectious disease	1.40 (0.85, 2.30)	0.182
Oncology	1.54 (1.05, 2.25)	0.028
Nephrology and urology	1.16 (0.56, 2.44)	0.687
Rheumatology	1.24 (0.25, 6.19)	0.793
Pulmonary medicine	2.27 (1.17, 4.42)	0.016
Orthopedics	2.27 (1.00, 5.17)	0.051
Psychiatry	2.67 (0.78, 9.12)	0.117
Ear-nose-throat	0.89 (0.33, 2.38)	0.818
Neurology	1.52 (0.69, 3.39)	0.301
Pediatrics	2.66 (0.91, 7.72)	0.073
Histology	1.06 (0.67, 1.68)	0.800
Clinical exam	1.04 (0.67, 1.62)	0.872
Imaging	0.98 (0.71, 1.34)	0.891
Biomarker	1.37 (0.96, 1.97)	0.087
Clinical challenge tests	2.88 (1.00, 8.33)	0.051
Physiologic tests	0.75 (0.39, 1.47)	0.409
Endoscopy	1.03 (0.40, 2.70)	0.948
Exact search described	1.70 (1.16, 2.50)	0.007
Search terms provided	1.17 (0.85, 1.60)	0.333
Search on demand (on non-journal website or from the authors)	1.39 (0.67, 2.86)	0.379
Years searched were reported	1.93 (1.14, 3.27)	0.014
Quality criteria for study selection	1.52 (1.00, 2.33)	0.052
Consideration of English-language studies only	0.63 (0.46, 0.87)	0.005
Consideration of at least one language other than English	1.70 (1.23, 2.36)	0.001
Medline	4.74 (2.23, 10.08)	<0.001
EMBASE	9.13 (6.04, 13.79)	<0.001
Conference proceedings	1.57 (0.92, 2.68)	0.097
Reference lists of eligible studies	1.20 (0.84, 1.71)	0.328
Reference lists of relevant review articles	1.22 (0.86, 1.74)	0.270
SCI or other ISI databases	5.21 (2.23, 12.19)	<0.001
Current Contents	0.37 (0.19, 0.73)	0.004
Experts in the field	0.79 (0.53, 1.18)	0.253
Test manufacturers	1.90 (0.77, 4.67)	0.161
Other specific electronic databases	6.80 (3.97, 11.66)	<0.001
Hand searching of journals	0.93 (0.60, 1.46)	0.759
Cochrane databases	16.25 (8.16, 32.35)	<0.001
CINAHL	12.56 (3.92, 40.24)	<0.001
At least one electronic database in addition to Medline	8.76 (6.13, 12.52)	<0.001
Unpublished information	0.72 (0.46, 1.14)	0.157
Any quality assessment	11.27 (7.81, 16.27)	<0.001
Study settings	1.21 (0.88, 1.68)	0.241
Consecutive patient recruitment	1.97 (1.41, 2.74)	<0.001

Prospective study design	3.11 (2.25, 4.30)	<0.001
Details of the reference standard test	0.68 (0.40, 1.13)	0.135
Reader expertise	2.72 (1.58, 4.67)	<0.001
Blinding (index test assessor to reference standard results)	NA	NA
Blinding (reference standard test assessor to index test results)	NA	NA
Blinding (unspecified)	NA	NA
Any blinding	2.32 (1.68, 3.20)	<0.001
Patient age	3.88 (2.74, 5.49)	<0.001
Patient sex	NA	NA
Location of primary studies	7.22 (4.08, 12.78)	<0.001
Spectrum bias	NA	NA
Selection bias	NA	NA
Time between index and reference standard test application	NA	NA
Test independence	NA	NA
Indeterminate test results	NA	NA
Withdrawals	NA	NA
Verification bias	4.33 (3.05, 6.15)	<0.001
QUADAS	NA	NA
STARD	NA	NA
Random effects used	5.21 (3.70, 7.32)	<0.001
Bayesian analyses	3.15 (0.72, 13.91)	0.129
OR	3.32 (2.25, 4.91)	<0.001
Sensitivity	1.77 (1.25, 2.53)	0.001
Specificity	2.02 (1.44, 2.84)	<0.001
Likelihood ratios	2.82 (1.91, 4.17)	<0.001
Accuracy	0.65 (0.34, 1.24)	0.196
Predictive values	0.91 (0.57, 1.45)	0.691
Q*	0.24 (0.11, 0.55)	0.001
AUC	0.60 (0.31, 1.15)	0.127
Any graphical display of analysis (synthesis) results	2.92 (2.08, 4.11)	<0.001
Forest plots	7.44 (4.81, 11.51)	<0.001
ROC space plots	1.66 (1.21, 2.27)	0.002
Heterogeneity testing	5.34 (3.80, 7.52)	<0.001
Univariate analyses	0.06 (0.02, 0.20)	<0.001
Comparative analyses	0.79 (0.53, 1.18)	0.253
Advanced synthesis methods	9.10 (3.29, 25.19)	<0.001
Any exploration of heterogeneity	1.89 (1.38, 2.59)	<0.001
Provides data for re-analyses	0.74 (0.54, 1.03)	0.071

AUC = area under the curve; CINAHL = Cumulative Index to Nursing and Allied Health Literature; NA = not available (indicates that estimation was not possible due to unavailability of data for a variable or perfect prediction); OR = odds ratio; QUADAS = Quality Assessment of Diagnostic Accuracy Studies; ROC = receiver operating characteristic; STARD = Standards for Reporting of Diagnostic Accuracy.

Appendix F. Regression Analyses for Citation Count and Journal Impact Factor

Appendix Table F1. Significant results from univariable regression analyses predicting the number of citations received by each paper per-year (all p-values were <0.001)

Factor	Relative Citation Rate (95% CI)
Clinical exam as test type	0.64 (0.54, 0.76)
Graphical presentation of results	1.39 (1.20, 1.61)
Number of studies	1.01 (1.00, 1.01)
Pediatrics as topic	0.52 (0.38, 0.72)
Use of advanced statistical methods	1.48 (1.21, 1.79)
Orthopedics as topic	0.62 (0.48, 0.81)
Reporting of comparative analyses between index tests	1.32 (1.13, 1.54)

CI = confidence interval

Appendix Table F2. Significant results from univariable regression analyses predicting meta-analysis methods or reporting characteristics using journal impact factor as a predictor (all p-values were <0.001)

Methods or Reporting Characteristic	OR (95% CI) per Impact-Factor Unit
Clinical exam as test type	1.06 (1.03, 1.08)
Use of likelihood ratios in quantitative analyses	1.05 (1.03, 1.08)
Use of random effects models	1.05 (1.03, 1.08)
Availability of the search strategy upon request	1.07 (1.03, 1.11)
Assessment of blinding	1.06 (1.02, 1.09)
Assessment of whether consecutive patients were enrolled	1.04 (1.02, 1.07)
Use of review papers to identify eligible studies for meta-analysis	1.04 (1.02, 1.07)

CI = confidence interval; OR = odds ratio

Appendix Table F3. Significant results from univariable regression analyses predicting meta-analysis methods or reporting characteristics using journal group as a predictor (high impact factor general medical journals versus all others; all p-values were <0.001)

Methods or Reporting Characteristic	OR (95% CI) (high-impact factor general medical journals versus all others)
Clinical exam as test type	3.75 (2.03, 6.92)
Use of random effects models	3.56 (1.78, 7.11)

CI = confidence interval; OR = odds ratio