JAMA | Special Communication

Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies
The PRISMA-DTA Statement

Matthew D. F. McInnes, MD; David Moher, PhD; Brett D. Thombs, PhD; Trevor A. McGrath, BSc; Patrick M. Bossuyt, PhD; and the PRISMA-DTA Group

IMPORTANCE Systematic reviews of diagnostic test accuracy synthesize data from primary diagnostic studies that have evaluated the accuracy of 1 or more index tests against a reference standard, provide estimates of test performance, allow comparisons of the accuracy of different tests, and facilitate the identification of sources of variability in test accuracy.

OBJECTIVE To develop the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagnostic test accuracy guideline as a stand-alone extension of the PRISMA statement. Modifications to the PRISMA statement reflect the specific requirements for reporting of systematic reviews and meta-analyses of diagnostic test accuracy studies and the abstracts for these reviews.

DESIGN Established standards from the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network were followed for the development of the guideline. The original PRISMA statement was used as a framework on which to modify and add items. A group of 24 multidisciplinary experts used a systematic review of articles on existing reporting guidelines and methods, a 3-round Delphi process, a consensus meeting, pilot testing, and iterative refinement to develop the PRISMA diagnostic test accuracy guideline. The final version of the PRISMA diagnostic test accuracy guideline checklist was approved by the group.

FINDINGS The systematic review (produced 64 items) and the Delphi process (provided feedback on 7 proposed items; 1 item was later split into 2 items) identified 71 potentially relevant items for consideration. The Delphi process reduced these to 60 items that were discussed at the consensus meeting. Following the meeting, pilot testing and iterative feedback were used to generate the 27-item PRISMA diagnostic test accuracy checklist. To reflect specific or optimal contemporary systematic review methods for diagnostic test accuracy, 8 of the 27 original PRISMA items were left unchanged, 17 were modified, 2 were added, and 2 were omitted.

CONCLUSIONS AND RELEVANCE The 27-item PRISMA diagnostic test accuracy checklist provides specific guidance for reporting of systematic reviews. The PRISMA diagnostic test accuracy guideline can facilitate the transparent reporting of reviews, and may assist in the evaluation of validity and applicability, enhance replicability of reviews, and make the results from systematic reviews of diagnostic test accuracy studies more useful.

Supplemental content

CME Quiz at jamanetwork.com/learning

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: The PRISMA-DTA Group authors and investigators appear at the end of the article.

Corresponding Author: Matthew D. F. McInnes, MD, Ottawa Hospital-Civic Campus, 1053 Carling Ave, Ottawa, ON K1E 4Y9, Canada (mmcinnes@toh.ca).

JAMA. 2018;319(4):388-396. doi:10.1001/jama.2017.19163

Systematic reviews can advance the understanding of diagnostic test accuracy. Systematic reviews of diagnostic test accuracy synthesize data from primary studies to provide insight into the ability of medical tests to detect a target condition; they also can provide estimates of test performance, allow comparisons of the accuracy of different tests, and facilitate the identification of sources of variability.[1] The number of systematic reviews of diagnostic test accuracy studies has increased rapidly; however, they are often not reported completely, which has contributed to "a crisis of repeatability."[2-5]

Reporting of systematic reviews should be complete and informative to enable readers to assess the quality of methods and the validity of the findings. Published systematic reviews of diagnostic test accuracy often have been uninformative and of heterogeneous quality.[4,6,7] They demonstrate variability in approaches to fundamental methodological steps, including methods to assess risk of bias, assessment of between-study variability, and methods for combining data across studies.[7-11]

To improve the reporting of systematic reviews, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guideline was published, which contained a 27-item checklist and flow diagram.[12] The initial PRISMA guideline was focused on improving the quality of systematic reviews of intervention studies; the authors of the original PRISMA statement suggested modification for diagnostic test accuracy reviews.[13] Although systematic reviews of diagnostic test accuracy studies share elements with those of intervention studies, there are important differences. Study design and measures of effect differ from those of randomized clinical trials. Accuracy can differ between studies due to differences in patients, setting, prior testing, and use of different reference standards. Consequently, the methods for evaluating risk of bias, summarizing results, and exploring variability for diagnostic test accuracy studies differ from those used for intervention studies. As such, some PRISMA items are not appropriate for systematic reviews of diagnostic test accuracy studies, others need adaptation, and some areas may not be covered.[1,14,15]

We aimed to develop the PRISMA diagnostic test accuracy guideline as a stand-alone extension of the PRISMA statement, modified to reflect the particular requirements for the reporting of diagnostic test accuracy studies in systematic reviews. A secondary objective was to identify items that should be included in the abstracts of systematic reviews of diagnostic test accuracy studies.

Key Points

Question What items should be reported to allow readers to evaluate the validity and applicability and to enhance the replicability of systematic reviews of diagnostic test accuracy studies?

Findings This diagnostic test accuracy guideline is an extension of the original Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. Two PRISMA items have been omitted, 2 were added, and 17 were modified to reflect specific or optimal contemporary systematic review methods of diagnostic test accuracy studies.

Meaning The guideline checklist can facilitate transparent reporting of reviews of diagnostic test accuracy studies, and may help assist evaluations of validity and applicability, enhance replicability of reviews, and make the results more useful for clinicians, journal editors, reviewers, guideline authors, and funders.

Methods

After establishing the PRISMA diagnostic test accuracy executive group, which was composed of the lead author of the PRISMA statement (D.M.),[12] the lead author of the Standards for the Reporting of Diagnostic Accuracy Studies (STARD) (P.M.B.),[16] and an experienced author, reviewer, and editor of systematic reviews of diagnostic test accuracy studies (M.D.F.M.), a number of experts were contacted to join the PRISMA diagnostic test accuracy group and assist with the project (all contacted experts agreed to participate). The goal was to assemble a team of experts in diagnostic test accuracy research and systematic review methods, complemented by authors, journal editors, funders, and users of systematic reviews of diagnostic test accuracy studies. The 24 members and their relevant expertise appear in eTable 1 in the Supplement.

The PRISMA diagnostic test accuracy executive group registered the protocol. Established standards from the Enhancing the Quality and Transparency of Health Research (EQUATOR) Network[17] were followed in the development of the guideline; no major deviations from the protocol occurred.[18] The PRISMA diagnostic test accuracy group used the original PRISMA statement[12] as a starting point and endeavored to identify items that needed to be added, removed, or modified to improve systematic reviews of diagnostic test accuracy studies.

Details of the systematic review for item generation have been published elsewhere.[19] To identify articles pertaining to the methods or reporting quality of systematic reviews of diagnostic test accuracy studies, searches of multiple databases and existing sources of guidance (eg, PRISMA, STARD 2015)[12,16] were performed. After performing data extraction from these reports, potential PRISMA diagnostic test accuracy items were categorized according to specific reporting topics: general overview, quality of reporting, search, variability, pooling methods, publication bias, risk of bias, and other. This list of potential PRISMA diagnostic test accuracy items was presented during the first round of the Delphi process.

Delphi Process

A 3-round Delphi process was held between December 2016 and March 2017 in which all members of the PRISMA diagnostic test accuracy group were invited to participate.[20,21] This modified Delphi process has been used previously for similar work such as Risk of Bias in Systematic Reviews and STARD 2015.[22,23] The aim of the process was to achieve consensus on essential items that should be included in the PRISMA diagnostic test accuracy guideline and to identify items that required discussion at the consensus meeting.

During each round of the survey process, potential essential items were proposed, and participants were asked to score each item on a Likert scale anchored at (1) "not essential to report in a systematic review of diagnostic test accuracy studies" and (5) "essential to report in a systematic review of diagnostic test

accuracy studies." Likert scores of 1 to 2 were categorized as a low score (item should not be part of PRISMA diagnostic test accuracy guideline); 3, moderate (item should be discussed), 4 to 5, high score (item should be part of PRISMA diagnostic test accuracy guideline). For an item to meet consensus, more than 66% of the Delphi respondents (>15 of 23) needed to rate 1 of these 3 categories; this threshold was based on what was used for previous reporting guidelines.[24]

During round 1 of the Delphi process, all items identified during the systematic review step were proposed.[19] Participants were also asked to suggest any additional items that were potentially relevant to report in systematic reviews of diagnostic test accuracy studies. Round 2 of the survey included any items that did not reach consensus during round 1, and any new items suggested by at least 1 respondent during round 1. As with round 2, round 3 involved items that did not reach consensus during rounds 1 or 2.

Following each of the 3 rounds, the mode (most frequent) score for each item was tabulated. Items were categorized as follows: (1) mode score of 1 to 3 but for less than 66% of participants, proceed to next round of Delphi process (or to a meeting discussion if this occurred during round 3 of the Delphi process); (2) consensus score of 1 or 2, do not include; (3) consensus score of 3, discuss at meeting; (4) mode score of 4 or 5 but for less than 66% of participants, discuss at meeting; and (5) consensus score of 4 or 5, include in PRISMA diagnostic test accuracy guideline (but discuss at meeting to confirm exact wording). All participants were provided an anonymized summary of the results after each round of the process. The survey was administered by SurveyMonkey Inc.

Consensus Meeting

A 2-day consensus meeting was held in Amsterdam, the Netherlands, in May 2017 and all members of the executive and PRISMA diagnostic test accuracy group were invited to attend. The main objective of this meeting was to agree on items for which no consensus was reached during the Delphi survey process and to generate a preliminary PRISMA diagnostic test accuracy checklist guideline (and a guideline for abstracts). For the items that reached consensus for inclusion prior to the meeting, the precise wording of the items was decided.

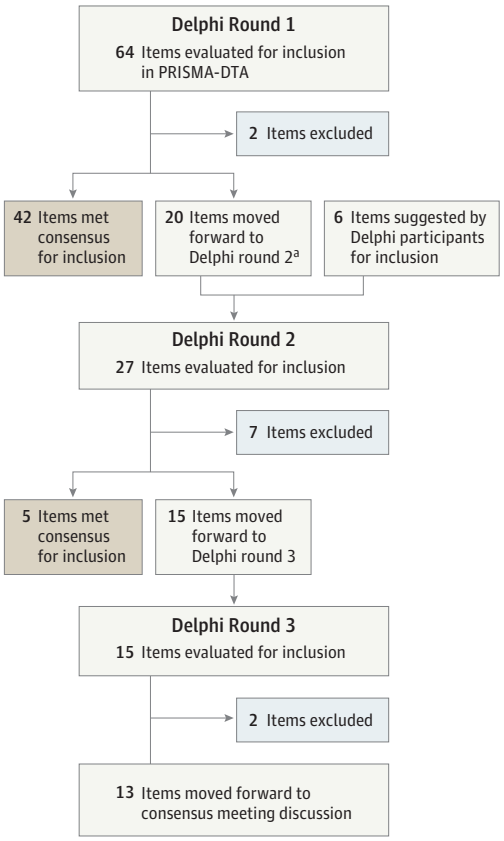
Checklist Pilot

Following the meeting, members of the PRISMA diagnostic test accuracy group reviewed and applied the checklist to ongoing systematic reviews of diagnostic test accuracy studies to identify any practical challenges with any of the items and to inform the writing of the statement. This included formal pilot testing by a graduate student (J-P.S.) of the preliminary checklist used to assess published systematic reviews of diagnostic test accuracy studies.

In addition, multiple potential users and interested parties (such as authors of systematic reviews and attendees of an author training course on conducting systematic reviews of diagnostic test accuracy studies) were invited to review and apply the preliminary checklist to assess utility and clarity of wording. Feedback from these pilot exercises was used to refine wording and presentation of the checklist. Formal feedback was gathered via a survey administered via SurveyMonkey, which was sent to the entire PRISMA diagnostic test accuracy group. Additional feedback was gathered via email correspondence. All sources of feedback were used to modify and inform the final version of the PRISMA diagnostic test accuracy checklist.

A further explanation and elaboration document will subsequently be developed to provide additional detail regarding the rationale for the items and examples. Based on government and institutional guidelines, this type of study does not require research ethics board approval.

Figure. Study Flow Diagram Documenting the Delphi Process



DTA indicates diagnostic test accuracy; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

[a] One item from round 1 was split into 2 items for round 2.

Results

Delphi Process

Twenty-three of 23 individuals (100%) completed all 3 rounds of the Delphi process (participation is documented in eTable 1 in the Supplement). During round 1, the group evaluated 64 items identified by the systematic review (**Figure**). Forty-two items met consensus for inclusion, 20 items were moved forward to round 2, 2 items were excluded, and an additional 6 items were suggested for inclusion for round 2.

During round 2, the group assessed 27 items (1 item from round 1 was split into 2). There were 5 items that met consensus for inclusion, 15 items were moved forward to round 3, and 7 items were excluded. During round 3, no items met consensus for

inclusion, 13 items were moved forward to the consensus meeting, and 2 items were excluded. Overall, after 3 Delphi rounds, 47 items were included (final wording to be discussed at the face-to-face consensus meeting), 13 items were moved forward to the consensus meeting to discuss inclusion or exclusion, and 11 items were excluded.

A list of the 11 excluded items appears in eTable 2 in the Supplement. Even though these items are considered relevant to the reporting of systematic reviews of diagnostic test accuracy studies, they were considered to be either too detailed for a minimum reporting guideline or not relevant depending on the scope or purpose of the review. Several of these items will be discussed further in the forthcoming explanation and elaboration document.

Consensus Meeting

Meeting attendance (n = 18) and the agenda are documented in eTables 1 and 3 in the Supplement, respectively. Of the 60 items discussed at the meeting, 27 were excluded. Excluded items and the rationale for exclusion are provided in eTable 2 in the Supplement.

Items 15 and 22 from the original 27-item PRISMA checklist were confirmed for removal. These items refer to the evaluation and reporting of risk of bias that may affect the cumulative evidence such as publication bias and selective reporting within studies. They were excluded for 2 main reasons. First, there is only limited evidence that publication or reporting bias is a major issue for primary diagnostic test accuracy studies.[25,26] As such, the rationale for mandating evaluation of bias in systematic reviews of diagnostic test accuracy studies is not as strong as for reviews of intervention studies. Second, there is no appropriate test with adequate statistical power to reliably assess publication bias in the context of diagnostic test accuracy systematic reviews.[27-29]

The remaining 33 items were discussed and synthesized into a draft checklist for PRISMA diagnostic test accuracy. Many of the items were combined to reduce redundancy between items and to minimize the total number of items. The PRISMA flow diagram was also reviewed at the consensus meeting and no modifications for PRISMA diagnostic test accuracy were deemed necessary.

Compared with the original PRISMA checklist, 2 new items were added. The first, labeled item D1, regards the statement of the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative reviews). The rationale for inclusion is 2-fold. First, the role of the index test is critical to understanding the place of a test in the diagnostic pathway; diagnostic accuracy can vary importantly depending on the clinical scenario. Without this information, generalizability of the results to the clinical setting may be limited.[16,30] Second, identifying minimally acceptable test accuracy may be helpful in forming conclusions. Whether a test is considered clinically useful cannot be determined by a diagnostic accuracy measure alone; its accuracy relative to alternative tests or management strategies must be considered, as well as the downstream consequences of false-positive and false-negative results. As such, considering external evidence to form criteria for minimally acceptable test accuracy standards may play an important role in forming the purpose of diagnostic test accuracy systematic reviews.[16,30,31]

Defining minimally acceptable test accuracy (or minimum difference) may not always be appropriate depending on the review question. For example, if a test is not yet well established or understood, the purpose of the review might be to evaluate reasons for variability in accuracy. For this reason, we have added the qualifier if applicable to this item.

The second new item is labeled item D2 and regards the reporting of the statistical methods used for the meta-analyses if performed. Meta-analyses of diagnostic test accuracy studies typically require multivariate models (eg, bivariate and hierarchical summary receiver operating characteristic), which allow for the tradeoff between sensitivity and specificity due to the positivity threshold, for potential correlation between estimates of sensitivity and specificity across studies, and for variability through the inclusion of random effects.[32,33] Traditional univariate methods ignore this correlation and can give misleading results.[5,34,35] We acknowledge that there are instances when univariate methods may be appropriate (eg, if the specificity of a test is set at 100%, or if the focus of the review is univariate meta-analysis of sensitivity). As such, reporting the method used for meta-analysis (if done) was considered essential for systematic reviews of diagnostic test accuracy studies.

Eight of the original PRISMA items (3, 5, 7, 9, 10, 16, 17, and 27) were not modified because they were considered to be equally applicable to systematic reviews of diagnostic test accuracy studies. Seventeen of the original PRISMA items (1, 2, 4, 6, 8, 11-14, 18-21, and 23-26) were adapted. The reasons for modification varied. The 2 major reasons were (1) there was unclear or ambiguous wording in the original PRISMA statement that required updating and (2) modified wording was necessary due to specific issues for systematic reviews of diagnostic test accuracy studies. **Table 1** lists the rationale for modification of the original PRISMA items for systematic reviews of diagnostic test accuracy. Further explanation and elaboration on the rationale and evidence will be provided in the forthcoming explanation and elaboration document.

At the consensus meeting, the original PRISMA checklist for abstracts was modified for systematic reviews of diagnostic test accuracy studies.[38] The total number of items was preserved (n = 12). Five items were not modified (4, 6, 10-12). One item was deleted (8, which was a description of the effect) because effect size is only relevant to intervention studies.[1,27] One new item was added (labeled A1, which regards synthesis of results) and corresponds to new item D2 in the PRISMA diagnostic test accuracy checklist. Six items were modified (1-3, 5, 7, and 9) to reflect the modified language for the corresponding items in the PRISMA diagnostic test accuracy checklist.

Pilot Testing and Revision

Thirty-seven points of feedback from the pilot exercise were received via email and formal survey. This feedback was considered by the PRISMA diagnostic test accuracy executive group and used to modify 5 of the items and add further explanation and rationale.

Final Checklists

The final version of the PRISMA diagnostic test accuracy checklist appears in **Table 2**. The new checklist has the same number of items as the original PRISMA checklist because 2 items were deleted (items 15 and 22) and 2 items were added(items D1 and D2); therefore, the numbering from the original PRISMA statement is preserved.

Table 1. Rationale for Modification of the Original Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement for Diagnostic Test Accuracy (DTA) Studies[a]

Section and Topic	Item No.	Reason for Modification
Title		
Title	1	Specification that the systematic review pertains to diagnostic test accuracy in the title is considered to enhance clarity of purpose and allow for easy identification when searching for reviews.
Abstract		
Structured summary	2	As per our study objectives, we aimed to create a specific, essential list of items for systematic reviews of DTA studies to be reported in the abstract. As such, we replaced this item with the PRISMA-DTA checklist for abstracts.[b]
Introduction		
Objectives	4	The original PRISMA wording (participant, intervention, comparison, outcome; PICO) was intended for systematic reviews of intervention studies. As such, the wording was modified to be more relevant for systematic reviews of DTA studies (eg, index test rather than intervention).[1,27]
Methods		
Eligibility criteria	6	Language specific to DTA (modifying PICO as described for item 4) was added.
Search	8	The primary reason for modification is not specific to DTA, but is for all contemporary systematic reviews. When the original PRISMA statement was written, it was not feasible to publish all electronic strategies. Present-day options for online supplemental material and institutional repositories provide options to report all search strategies, which will enhance transparency, improve replicability, and enable easier updating of systematic reviews.
Data items	11	Data items and relevant definitions with language specific to and essential regarding study objectives and risk of bias in systematic reviews of DTA studies (index test, target condition) were modified.[15,16,30,36]
Risk of bias in individual studies	12	Individual DTA studies may be at risk of bias and there can also be concerns regarding applicability (as highlighted in QUADAS-2). As such, language to reflect this was added.[15]
Summary measures	13	Summary measures: 1. The measures provided in PRISMA (eg, risk ratios) are specific to systematic reviews of intervention effectiveness. As such, the wording was modified to reflect measures relevant to assessing diagnostic accuracy (eg, sensitivity).[16,30] 2. The unit of assessment (per lesion with multiple liver lesion samples treated as individual observations in a 2 × 2 table vs per patient) can be critical in regard to accuracy estimates and generalizability of results due to potential bias introduced from clustering effects in per-lesion analysis.[37] As such, this additional requirement relevant to systematic reviews of DTA studies was added.
Synthesis of results	14	Synthesis of results 1. Measures of consistency (eg, I^2) considered routine in reviews of intervention studies are not typically applicable in systematic reviews of DTA studies. There is no consensus regarding alternative statistics. As such, the more general term *variability*, which can reflect multiple strategies to explore variability, was used in place of the term *inconsistency*.[8,10,14] 2. Additional specific items particular to systematic reviews of DTA studies were considered to be of sufficient relevance to list as requirements. These include describing definitions of the target condition, test positivity, and others.[16,30]
Study characteristics	18	Study characteristics considered to be essential regarding risk of bias and applicability in systematic reviews of DTA studies were listed (eg, reference standard, clinical setting).[27] The category of funding sources was added to optimize transparency because industry vs nonindustry funding may be relevant to consider in systematic reviews of DTA studies.
Risk of bias within studies	19	Individual DTA studies may be at risk of bias and there also can be concerns regarding applicability (as highlighted in QUADAS-2). As such, language to reflect this was added.[15]
Results		
Results of individual studies	20	Original wording for summary data for each intervention group was specific to systematic reviews of intervention effectiveness. As such, the wording was revised to reflect results relevant to systematic reviews of DTA studies (eg, 2 × 2 data and positivity threshold used).[15,16,30] Reporting of 2 × 2 data is required to allow readers to evaluate important variables such as the proportion with the target condition and other accuracy estimates that may not have been specifically addressed in the review (eg, positive predictive value).
Synthesis of results	21	Language was modified to be more specific and relevant to systematic reviews of DTA studies (eg, describe test accuracy). In addition, the term *inconsistency* was replaced with the preferred term *variability* (as described in item 14) for systematic reviews of DTA studies.
Additional analyses	23	In addition to the original PRISMA wording, we ask that additional information (including potential harms) relevant to systematic reviews of DTA studies be reported (eg, index test failures such as inconclusive, unusable, or indeterminate results and adverse events related to index test administration).[16,30]
Summary of evidence	24	Wording was simplified to only refer to main findings because there is typically only 1 primary outcome (diagnostic accuracy) in systematic reviews of DTA studies. In addition, relevance to key groups was considered to be more appropriate with item 26; as such, it was modified and moved to item 26.
Discussion		
Limitations	25	As per discussion for item 12, wording was modified to reflect concerns regarding the term *applicability* in addition to the term *risk of bias*.[15]
Conclusions	26	As per discussion for item 24, implications for clinical practice were considered to be more appropriate in the conclusions. In addition, language specific to the generalizability of the findings for systematic reviews of DTA studies (eg, intended use and clinical role of the index test) was added.[16,30]

Abbreviation: QUADAS-2, second version of Quality Assessment of Diagnostic Accuracy Studies.

[a] The final wording for the PRISMA-DTA checklist appears in Table 2 and the abstract checklist appears in Table 3.

[b] The checklist for abstracts appears in Table 3.

The final version of the PRISMA diagnostic test accuracy checklist for abstracts appears in **Table 3** and has the same number of items as the original PRISMA checklist for abstracts because 1 item was deleted (item 8) and 1 item was added (item A1); therefore, the numbering from the original PRISMA for abstracts is preserved.

Table 2. Checklist for Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for Diagnostic Test Accuracy (DTA) Studies

Section and Topic	Item No.[a]	Description
Title		
Title[b]	1	Identify the report as a systematic review (meta-analysis) of DTA studies.
Abstract		
Abstract[b]	2	A checklist for abstracts appears in Table 3.
Introduction		
Rationale[c]	3	Describe the rationale for the review in the context of what is already known
Clinical role of index test[d]	D1	State the scientific and clinical background, including the intended use and clinical role of the index test, and if applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for a comparative design).
Objectives[b]	4	Provide an explicit statement of question being addressed in terms of participants, index test, and target conditions.
Methods		
Protocol and registration[c]	5	Indicate where the review protocol can be accessed (eg, web address) and provide trial registration number if available.
Eligibility criteria[b]	6	Specify study characteristics (participants, setting, index test, reference standards, target conditions, and study design) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility and providing rationale.
Information sources[c]	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and the date last searched.
Search[b]	8	Present full search strategies for all electronic databases and other sources searched, including any limits used so that they can be repeated.
Study selection[c]	9	State the process for selecting studies (ie, screening, eligibility, whether included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process[c]	10	Describe the methods of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from the investigators.
Definitions for data extraction[b]	11	Provide definitions used in data extraction and classifications of target conditions, index tests, reference standards, and other characteristics (eg, study design, clinical setting).
Risk of bias and applicability[b]	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.
Diagnostic accuracy measures[b]	13	State the principal diagnostic accuracy measures reported (eg, sensitivity, specificity) and state the unit of assessment (eg, per patient vs per lesion).
Synthesis of results[b]	14	Describe the methods of handling the data, combining the results of the studies and describing the variability between studies. This could include, but is not limited to (1) handling of multiple definitions of the target condition, (2) handling of multiple thresholds of test positivity, (3) handling multiple index test readers, (4) handling of indeterminate test results, (5) grouping and comparing tests, and (6) handling of different reference standards.
Meta-analysis[d]	D2	Report the statistical methods used for meta-analyses if performed.
Additional analyses[c]	16	Describe the methods of the additional analyses (eg, sensitivity or subgroup analyses, meta-regression) if done, indicating which were prespecified.
Results		
Study selection[c]	17	Provide the numbers of studies screened, assessed for eligibility, included in the review, and included in the meta-analysis if applicable, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics[b]	18	For each included study, provide citations and present key characteristics including (1) participant characteristics (presentation, prior testing), (2) clinical setting, (3) study design, (4) target condition definition, (5) index test, (6) reference standard, (7) sample size, and (8) funding sources.
Risk of bias and applicability[b]	19	Present evaluation of risk of bias and concerns regarding applicability for each study.
Results of individual studies[b]	20	For each analysis in each study (eg, unique combination of index test, reference standard, and positivity threshold), report 2 × 2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest plot or a receiver operating characteristic curve.
Synthesis of results[b]	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.
Additional analyses[b]	23	Give results of additional analyses if done (eg, sensitivity or subgroup analyses, meta-regression, analysis of index test, failure rates, proportion of inconclusive results, and adverse events).
Discussion		
Summary[b]	24	Summarize the main findings including the strength of the evidence.
Limitations[b]	25	Discuss limitations from included studies (eg, risk of bias and concerns regarding applicability) and from the review process (eg, incomplete retrieval of identified research).
Conclusions[b]	26	Provide a general interpretation of the results in the context of other evidence. Discuss implications for future research and clinical practice (eg, the intended use and clinical role of the index test).
Other		
Funding[c]	27	For the systematic review, describe the sources of funding and other support and the role of the funders.

Abbreviations: FN, false negative; FP, false positive; TN, true negative; TP, true positive.

[a] Original PRISMA items 15 and 22 were omitted for reasons listed in the Consensus Meeting subsection in the Results section of the text.

[b] Modified original PRISMA item.

[c] Unmodified PRISMA item.

[d] New PRISMA-DTA item.

Table 3. Abstract Checklist for Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for Diagnostic Test Accuracy (DTA) Studies

Section and Topic	Item No.[a]	Description
Title and Purpose		
Title[b]	1	Identify the report as a systematic review (meta-analysis) of DTA studies.
Objectives[b]	2	Indicate the research question, including components such as participants, index test, and target conditions.
Methods		
Eligibility criteria[b]	3	Include study characteristics used as criteria for eligibility.
Information sources[c]	4	List the key databases searched and the search dates.
Risk of bias and applicability	5	Indicate the methods of assessing risk of bias and applicability.
Synthesis of results[d]	A1	Indicate the methods for the data synthesis.
Results		
Included studies[c]	6	Indicate the number and type of included studies and the participants and relevant characteristics of the studies (including the reference standard).
Synthesis of results[b]	7	Include the results for the analysis of diagnostic accuracy, preferably indicating the number of studies and participants. Describe test accuracy including variability; if meta-analysis was done, include summary results and confidence intervals.
Discussion		
Strengths and limitations[b]	9	Provide a brief summary of the strengths and limitations of the evidence.
Interpretation[c]	10	Provide a general interpretation of the results and the important implications.
Other		
Funding[c]	11	Indicate the primary source of funding for the review.
Registration[c]	12	Provide the registration number and the registry name

[a] Original PRISMA item 8 was omitted for reasons listed in the Consensus Meeting subsection in the Results section of the text.

[b] Modified original PRISMA item.

[c] Unmodified PRISMA item.

[d] New PRISMA-DTA item.

Discussion

The PRISMA diagnostic test accuracy checklist provides guidance specific to systematic reviews of diagnostic test accuracy studies. Both the PRISMA diagnostic test accuracy checklist and the checklist for abstracts were developed with multidisciplinary consensus approaches as per best practices for guideline development.[18] The PRISMA diagnostic test accuracy checklist items reflect the concepts, methods, and language specific to systematic reviews of diagnostic test accuracy studies and, if implemented, can help ensure that information for assessment of risk of bias and applicability and can enhance transparency and replicability of systematic reviews of diagnostic test accuracy studies. This work should be of practical use to those who author, review, publish, fund, and implement the results of systematic reviews of diagnostic test accuracy studies. It may also be useful as a guidance for protocols of systematic reviews on diagnostic test accuracy. This checklist is relevant to include for the evaluation of single tests, multiple tests (comparative), and multivariable diagnostic models.

The PRISMA diagnostic test accuracy checklist aims to improve the completeness and transparency of reporting of systematic reviews of diagnostic test accuracy studies. Complete reporting might be associated with review quality, however, they are not inseparable.[4] The understanding and the application of the optimal principles and methods for systematic reviews of diagnostic test accuracy studies are complex and require knowledge acquired from resources beyond a 27-item reporting checklist.[1] Even though guidance is available for conducting systematic reviews of diagnostic test accuracy studies,[27] considerable areas of uncertainty remain (eg, optimal methods for assessing variability, appropriate interpretation of review findings); these areas are likely to evolve based on ongoing and future research.[8] As such, prospective reviewers are encouraged to seek specialized training (eg, Cochrane group author training resources for screening and diagnostic test methods) and to collaborate with those experienced in systematic review methods for diagnostic test accuracy studies.[39]

Conforming to reporting guidelines can be challenging based on journal-level constraints such as limits on words, tables, and figures; however, there is little evidence to indicate that reporting guidelines increase the word count of articles. Methods to ensure complete reporting may include the use of supplementary online material, institutional repositories, and appendices. The PRISMA diagnostic test accuracy guideline represents minimum reporting requirements, rather than a constraint or cap on what should be reported. Additional information that authors consider relevant to their specific review question may also be reported (eg, interobserver agreement for imaging reviews).

Complete reporting of diagnostic test accuracy systematic reviews may be hindered by incomplete reporting in diagnostic test accuracy primary studies.[40] This challenge makes complete

reporting of systematic reviews of diagnostic test accuracy studies more important because readers need to know whether the necessary information from the primary studies was available and whether conclusions can be drawn based on that information.

Limitations

Development of the PRISMA diagnostic test accuracy statement was guided by evidence-based principles when possible; however, when evidence was lacking, we relied on expert opinion. The PRISMA diagnostic test accuracy checklist was designed for all types of diagnostic test accuracy research; some specialties (eg, imaging) may have important items unique to their specialty (eg, interobserver agreement) that were not included in the guideline but should be reported. In addition, as the body of evidence in diagnostic test accuracy research grows, the PRISMA diagnostic test accuracy guideline will need to be updated to reflect these advances.

Conclusions

The 27-item PRISMA diagnostic test accuracy checklist provides specific guidance for reporting of systematic reviews. The PRISMA diagnostic test accuracy guideline can facilitate the transparent reporting of reviews, and may assist in the evaluation of validity and applicability, enhance replicability of reviews, and make the results from systematic reviews of diagnostic test accuracy studies more useful.

ARTICLE INFORMATION

Accepted for Publication: December 6, 2017.

The PRISMA-DTA Group Authors: Tammy Clifford, PhD; Jérémie F. Cohen, MD, PhD; Jonathan J. Deeks, PhD; Constantine Gatsonis, PhD; Lotty Hooft, PhD; Harriet A. Hunt, MSc; Christopher J. Hyde, PhD; Daniël A. Korevaar, MD, PhD; Mariska M. G. Leeflang, PhD; Petra Macaskill, PhD; Johannes B. Reitsma, MD, PhD; Rachel Rodin, MD, MPH; Anne W. S. Rutjes, PhD; Jean-Paul Salameh, BSc; Adrienne Stevens, MSc; Yemisi Takwoingi, PhD; Marcello Tonelli, MD, SM; Laura Weeks, PhD; Penny Whiting, PhD; Brian H. Willis, MD, PhD.

Affiliations of The PRISMA-DTA Group Authors: Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada (Salameh); Department of Clinical Epidemiology, Biostatistics and Bioinformatics, University of Amsterdam, Academic Medical Center, Amsterdam, the Netherlands (Korevaar, Leeflang); Canadian Agency for Drugs and Technologies in Health, Ottawa, Ontario (Clifford, Weeks); Department of Pediatrics, Necker-Enfants Malades Hospital, Assistance Publique Hôpitaux de Paris, Paris Descartes University, Paris, France (Cohen); Inserm UMR 1153, Research Center for Epidemiology and Biostatistics Sorbonne Paris Cité, Paris Descartes University, Paris, France (Cohen); University of Birmingham, Birmingham, England (Deeks, Takwoingi, Willis); Brown University, Providence, Rhode Island (Gatsonis); Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands (Hooft, Reitsma); University of Exeter, Exeter, England (Hunt, Hyde); University of Sydney, Sydney, Australia (Macaskill); Public Health Agency of Canada, Ottawa, Ontario, Canada (Rodin); Institute of Social and Preventive Medicine, Berner Institut für Hausarztmedizin, University of Bern, Bern, Switzerland (Rutjes); School of Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada (Salameh); Ottawa Hospital Research Institute, Ottawa, Ontario, Canada (Stevens); Translational Research in Biomedicine Program, School of Medicine, University of Split, Split, Croatia (Stevens); University of Calgary, Calgary, Alberta, Canada (Tonelli); University of Bristol, National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West, Bristol, England (Whiting).

Author Affiliations: Department of Radiology, University of Ottawa, Ottawa, Ontario, Canada (McInnes); Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada (McInnes, Moher); Lady Davis Institute of the Jewish General Hospital, Montreal, Quebec, Canada (Thombs); Department of Psychiatry, McGill University, Montreal, Quebec, Canada (Thombs); University of Ottawa Department of Radiology, Ottawa, Ontario, Canada (McGrath); Department of Clinical Epidemiology, Biostatistics and Bioinformatics, University of Amsterdam, Academic Medical Center, Amsterdam, the Netherlands (Bossuyt).

Author Contributions: Dr McInnes had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: McInnes, Moher, McGrath, Bossuyt, Clifford, Cohen, Korevaar, Reitsma, Salameh, Takwoingi, Willis.
Acquisition, analysis, or interpretation of data: McInnes, Thombs, McGrath, Cohen, Gatsonis, Hunt, Hyde, Korevaar, Leeflang, Reitsma, Rutjes, Salameh, Stevens, Takwoingi, Tonelli, Weeks, Whiting, Willis.
Drafting of the manuscript: McInnes, Moher, McGrath, Leeflang, Salameh, Willis.
Critical revision of the manuscript for important intellectual content: McInnes, Thombs, McGrath, Bossuyt, Clifford, Cohen, Deeks, Gatsonis, Hooft, Hunt, Hyde, Korevaar, Leeflang, Macaskill, Reitsma, Rodin, Rutjes, Salameh, Stevens, Takwoingi, Tonelli, Weeks, Whiting, Willis.
Obtained funding: McInnes, Clifford.
Administrative, technical, or material support: McInnes, Moher, Clifford, Hunt, Salameh, Willis.
Supervision: McInnes, Bossuyt, Takwoingi.

Conflict of Interest Disclosures: The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

Funding/Support: The research was supported by grant 375751 from the Canadian Institute for Health Research; funding from the Canadian Agency for Drugs and Technologies in Health; funding from the Standards for Reporting of Diagnostic Accuracy Studies Group; funding from the University of Ottawa Department of Radiology Research Stipend Program; and funding from the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care South West Peninsula.

Role of the Funder/Sponsor: None of the funding sources had any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. McInnes MD, Bossuyt PM. Pitfalls of systematic reviews and meta-analyses in imaging research. *Radiology*. 2015;277(1):13-21.

2. Bastian H, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med*. 2010;7(9):e1000326.

3. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.

4. Tunis AS, McInnes MD, Hanna R, Esmail K. Association of study quality with completeness of reporting: have completeness of reporting and quality of systematic reviews and meta-analyses in major radiology journals changed since publication of the PRISMA statement? *Radiology*. 2013;269(2):413-426.

5. McGrath TA, McInnes MD, Korevaar DA, Bossuyt PM. Meta-analyses of diagnostic accuracy in imaging journals: analysis of pooling techniques and their effect on summary estimates of diagnostic accuracy. *Radiology*. 2016;281(1):78-85.

6. Willis BH, Quigley M. The assessment of the quality of reporting of meta-analyses in diagnostic research: a systematic review. *BMC Med Res Methodol*. 2011;11:163.

7. 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.

8. Naaktgeboren CA, van Enst WA, Ochodo EA, et al. Systematic overview finds variation in approaches to investigating and reporting on sources of heterogeneity in systematic reviews of diagnostic studies. *J Clin Epidemiol*. 2014;67(11):1200-1209.

9. Ochodo EA, van Enst WA, Naaktgeboren CA, et al. Incorporating quality assessments of primary studies in the conclusions of diagnostic accuracy reviews: a cross-sectional study. *BMC Med Res Methodol*. 2014;14:33.

10. Naaktgeboren CA, Ochodo EA, Van Enst WA, et al. Assessing variability in results in systematic reviews of diagnostic studies. *BMC Med Res Methodol*. 2016;16(1):6.

11. McGrath TA, McInnes MDF, Langer FW, Hong J, Korevaar DA, Bossuyt PMM. Treatment of multiple test readers in diagnostic accuracy systematic reviews-meta-analyses of imaging studies. *Eur J Radiol*. 2017;93:59-64.

12. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006-1012.

13. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1-e34.

14. Macaskill PGC, Deeks JJ, Harbord RM, Takwoingi Y. Analysing and presenting results. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. Oxford, England: Cochrane Collaboration; 2010.

15. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.

16. Bossuyt PM, Reitsma JB, Bruns DE, et al; STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Radiology*. 2015;277(3):826-832.

17. Equator Network. Reporting guidelines under development. http://www.equator-network.org/library/reporting-guidelines-under-development/#99. Accessed November 28, 2014.

18. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.

19. McGrath TA, Alabousi M, Skidmore B, et al. Recommendations for reporting of systematic reviews and meta-analyses of diagnostic test accuracy: a systematic review. *Syst Rev*. 2017;6(1):194.

20. Trevelyan E, Robinson N. Delphi methodology in health research: how to do it? *Eur J Integr Med*. 2015;7:423-428.

21. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS One*. 2011;6(6):e20476.

22. Whiting P, Savović J, Higgins JP, et al; ROBIS Group. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol*. 2016;69:225-234.

23. Korevaar DA, Cohen JF, Reitsma JB, et al. Updating standards for reporting diagnostic accuracy: the development of STARD 2015 [published online June 7, 2016]. *Res Integr Peer Rev*. doi:10.1186/s41073-016-0014-7

24. Cohen JF, Korevaar DA, Gatsonis CA, et al; STARD Group. STARD for abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ*. 2017;358:j3751.

25. Korevaar DA, Bossuyt PM, Hooft L. Infrequent and incomplete registration of test accuracy studies: analysis of recent study reports. *BMJ Open*. 2014;4(1):e004596.

26. Korevaar DA, Cohen JF, Spijker R, et al. Reported estimates of diagnostic accuracy in ophthalmology conference abstracts were not associated with full-text publication. *J Clin Epidemiol*. 2016;79:96-103.

27. Deeks J, Bossuyt P, Gatsonis C. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. 1.0.0 ed. Oxford, England: Cochrane Collaboration; 2013.

28. van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol*. 2014;14:70.

29. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-893.

30. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.

31. McGrath TA, McInnes MDF, van Es N, Leeflang MMG, Korevaar DA, Bossuyt PMM. Overinterpretation of research findings: evidence of "spin" in systematic reviews of diagnostic accuracy studies. *Clin Chem*. 2017;63(8):1353-1362.

32. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001;20(19):2865-2884.

33. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.

34. Dinnes J, Mallett S, Hopewell S, Roderick PJ, Deeks JJ. The Moses-Littenberg meta-analytical method generates systematic differences in test accuracy compared to hierarchical meta-analytical models. *J Clin Epidemiol*. 2016;80:77-87.

35. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. *J Clin Epidemiol*. 1995;48(1):119-130.

36. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061-1066.

37. Zwinderman AH, Glas AS, Bossuyt PM, Florie J, Bipat S, Stoker J. Statistical models for quantifying diagnostic accuracy with multiple lesions per patient. *Biostatistics*. 2008;9(3):513-522.

38. Beller EM, Glasziou PP, Altman DG, et al; PRISMA for Abstracts Group. PRISMA for abstracts: reporting systematic reviews in journal and conference abstracts. *PLoS Med*. 2013;10(4):e1001419.

39. de Vet HCW, Eisinga A, Riphagen II, Aertgeerts B, Pewsner D. Searching for studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0.4*. Oxford, England: Cochrane Collaboration; 2008.

40. Hong PJ, Korevaar DA, McGrath TA, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: adherence to STARD 2015 [published online June 22, 2017]. *J Magn Reson Imaging*. doi:10.1002/jmri.25797