Title: Completeness of reporting of systematic reviews of diagnostic test accuracy based on the PRISMA-DTA reporting guideline

Authors:

1. Jean-Paul Salameh BSc. Clinical Epidemiology Program, Ottawa Hospital Research Institute. School of Epidemiology, Public Health and Preventative Medicine, Faculty of Medicine, University of Ottawa. jsala016@uottawa.ca
2. Matthew DF McInnes MD FRCPC (Corresponding Author). Associate Professor, University of Ottawa Department of Radiology. Clinical Epidemiology Program, Ottawa Hospital Research Institute. Room c159 Ottawa Hospital Civic Campus, 1053 Carling Ave. Ottawa ON, K1Y 4E9. Email: mmcinnes@toh.on.ca Phone: (613)761-4054. Fax: (613)761-4476
3. David Moher PhD. The Ottawa Hospital Research Institute Clinical Epidemiology Program (Centre for Journalology). dmoher@ohri.ca
4. Brett D. Thombs PhD, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Canada; Departments of Psychiatry; Medicine; Epidemiology, Biostatistics and Occupational Health; Psychology; and Educational and Counselling Psychology, McGill University, Montréal, Canada. brett.thombs@mcgill.ca
5. Trevor A McGrath BSc. Department of Radiology, Faculty of Medicine, University of Ottawa. tmcgr043@uottawa.ca
6. Robert Frank BSc, Department of Radiology, Faculty of Medicine, University of Ottawa. rfran048@uottawa.ca
7. Anahita Dehmoobad Sharifabadi BHSc, Department of Radiology, Faculty of Medicine, University of Ottawa. adehm071@uottawa.ca
8. Noémie Kraaijpoel, Department of Vascular Medicine, Academic Medical Center, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands. n.kraaijpoel@amc.uva.nl
9. Brooke Levis MSc, Lady Davis Institute for Medical Research, Jewish General Hospital, Montréal, Canada; Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montréal, Canada. brooke.levis@mail.mcgill.ca
10. Patrick M Bossuyt PhD, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam. p.m.bossuyt@amc.nl

Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

Funding

1. This study was funded by the Canadian Institute for Health Research (CIHR; Grant Number 375751)
2. Mr. Salameh was supported by the Ontario Graduate Scholarship (OGS) award
3. Dr. McInnes was supported by the University of Ottawa Department of Radiology Research Stipend Program
4. Dr. Thombs was supported by a Fonds de recherche du Québec - Santé (FRQS) researcher salary award
5. Ms. Levis was supported by a CIHR Frederick Banting and Charles Best Canada Graduate Scholarship doctoral award

Role of Funders:

None of the funding bodies listed had any role in: 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.

Key Words: Diagnostic Accuracy; Meta-Analysis; Reporting Guidelines

Word count: 2909

Abstract:

Objective: To evaluate the completeness of reporting of diagnostic test accuracy (DTA) systematic reviews using the recently developed PRISMA-DTA guidelines.

Methods: Medline was searched for DTA systematic reviews published October 2017-Janaury 2018. The search time span was modulated to reach the desired sample size of 100 systematic reviews. Reporting on a per item basis using PRISMA-DTA was evaluated. Associations between reporting completeness and journal/study-level variables were examined. Correlation of reporting completeness with word count(abstract/full-text) was assessed.

Results: 100 reviews were included. Mean reported items=18.6/26(71%,SD=1.9) for PRISMA-DTA; 5.5/11(50%,SD=1.2) for PRISMA-DTA for abstracts. Items in the results were frequently reported; items related to: protocol registration, characteristics of included studies, results-synthesis, and definitions used in data extraction were infrequently reported. Infrequently reported items from PRISMA-DTA for abstracts included funding information, strengths and limitations, characteristics of included studies, and assessment of applicability. Reporting completeness was higher in higher impact factor journals(18.9vs.18.1 items;*P*=0.04), studies that cited PRISMA(18.9vs.17.7 items;*P*=0.003) or used supplementary-material(19.1vs.18.0 items;*P*=0.004). Variability in reporting was associated with author country(*P*=0.04), but not journal(*P*=0.6), abstract word count limitations(*P*= 0.9), PRISMA-adoption(*P*=0.2), structured-abstracts(*P*=0.2), study-design(*P*=0.8), subspecialty area(*P*=0.09), nor index test(*P*=0.5). Abstracts with a higher word count were more informative (R=0.4; *P*<0.001) but no association with word counts was observed for full-text reports (R=-0.03;*P*=0.06).

Conclusion: Recently published reports of DTA systematic reviews are not fully informative, when evaluated against the PRISMA-DTA guidelines. These results should guide knowledge translation strategies including journal level (adoption of PRISMA-DTA, increased abstract word count and use of supplementary material) and author level (PRISMA-DTA citation-awareness) strategies.

Introduction

Improving our understanding of the performance of diagnostic tests was recently identified as a priority by the Institute of Medicine[1,2]. Systematic reviews of diagnostic test accuracy (DTA) research are increasingly common and require unique methodological approaches in order to optimize the validity of the results[3-8]. Although clinicians and policy makers often rely on systematic reviews as high-level evidence, many systematic reviews (DTA included) do not report all of the information necessary assess the validity and generalizability of results[9-11]. More informative reporting will allow the many stakeholders who rely on DTA systematic reviews (e.g., clinicians, journal editors, guidelines authors and funding agencies) to better assess critical aspects of review methods and quality of evidence in order to evaluate the applicability and validity of reviews to clinical settings.

Reporting guidelines are checklists (and often flow diagrams) specifying the minimum information that should be provided in an article to ensure high quality and completeness of reporting, prerequisites to any efforts of reproducibility. In order to improve the transparent reporting of systematic reviews, various reporting guidelines and checklists have been developed [9,12-17]. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement was published to help improve completeness of reporting for systematic reviews and consists of 27 items and a flow diagram[18]. Since the methodological approach of DTA studies differs notably from intervention studies [8,19], PRISMA-DTA (and PRISMA-DTA for abstracts) were recently published as extensions of the PRISMA statement for DTA systematic reviews to address these differences[20].

The current level of completeness of reporting of DTA systematic reviews is not known. An evaluation of the level of completeness and informativeness of reports of DTA systematic reviews, using the PRISMA DTA guidelines, could guide knowledge translation strategies aimed at improving reporting of these reviews, specifically targeting those items and features that are often poorly reported.

The purpose of this study is to evaluate the level of completeness of recently published DTA systematic reviews, using the PRISMA DTA and PRISMA-DTA for abstracts reporting guidelines, and to explore variables potentially associated with completeness.

Methods

The study protocol was registered in the Open Science Framework (DOI 10.17605/OSF.IO /JDQWN); no major protocol deviations occurred. Research ethics board approval was not required.

Search

MEDLINE was searched for DTA systematic reviews published between October 31st, 2017 and January 20th, 2018 using the following previously published search strategy[6]: systematic[sb] AND (sensitivity and specificity[mesh] OR sensitivit*[tw] OR specifit*[tw] OR accur*[tw] or ROC[tw] or AUC[tw] or likelihood[tw]). The time span of the search was modulated to reach the desired sample size of 100 systematic reviews, starting with the month of publication of the PRISMA-DTA document and including additional previous months until the desired sample size was reached[12]. Sample size was based on convenience, feasibility and other recent publications on reporting guideline completeness[9,12,21].

Article selection

Eligible articles were full reports of systematic reviews that had evaluated the diagnostic accuracy of one or more index tests on humans by comparing it against a reference standard. Reports not published in English were excluded. Initial screening of search results based on title and abstract was done by one reviewer (J.P.S. - graduate student), and decisions about inclusion based on full text were done independently by 2 reviewers (JPS and TM - medical student). Disagreements were discussed with MDFM (Radiologist/Scientist) and resolved by consensus.

Data Extraction

The following data from included articles were extracted by one author (JPS): First author surname, country of corresponding author's institution, journal, journal impact factor (2016 one-year impact factor), year of publication, subspecialty area, index-test type (e.g., laboratory, imaging), study design (single test vs. comparative), abstract word count limitation by journal (yes/no), structure of abstract (structured vs. unstructured), word count (abstract and full text excluding supplementary material), use of supplementary material (yes/no), journal PRISMA adoption (yes/no) and whether the study cited PRISMA (or a PRISMA extension). Six

extractors (JPS, all studies; AD, RF, and TM - medical students, NK - MD, and BL- PhD candidate, 20% of the studies each) independently assessed the overall completeness relative to the 26 PRISMA-DTA reporting requirements (full checklist of 27 items less the item referring to PRISMA-DTA for abstracts) as well as to the 11 PRISMA-DTA for abstracts reporting requirements for each included study. Each reporting requirement was rated as 'Yes', 'No' or 'N/A' with any disagreements resolved by consensus. Items were rated as 'N/A' when, for instance, no additional analyses were done (Item 22). 'N/A' items were treated as a 'Yes' during data analysis. Appendices 1 and 2 include the PRISMA-DTA and PRISMA-DTA for abstracts elements, respectively. If the item was reported anywhere in the article (or in the abstract for PRISMA-DTA for abstracts) it was scored as a 'yes', unless specified within the item description that it must be reported in a specific section (e.g., item 1 in the title/abstract). Information could have been included either in the full text report or in the supplementary material (including online-only material) to be rated as 'yes'. Instructions for authors for each included journal were assessed to determine whether the journal is a PRISMA adopter or not.

To optimize inter-observer agreement, two strategies were used: (1) a pilot extraction for 4 articles not included in the analysis was performed after a training session on the extraction process; (2) a 'user's guide' (Appendix 3) with descriptions of the rating process of specific items was created during the pilot exercise for reference during data extraction.

Data Analysis

The overall completeness of reporting, evaluated against the PRISMA-DTA guidelines (out of 26 items) and completeness on a per-item basis were calculated. Items with multiple sub-points (a, b, etc.) were scored with a total of 1 point with fractional points awarded for each sub-item (e.g., 0.5 points each if 2 sub-items).

Association of completeness of reporting with: journal, country, impact factor, index test type (e.g., imaging, laboratory), journal PRISMA adoption, citation of PRISMA (or extension), use of supplementary material, and word count (abstract word count for PRISMA-DTA for abstracts) was evaluated. A previously reported, descriptive classification of reporting was applied as follows: items reported in <33% of studies were considered "infrequently reported," those reported in 33–66% of studies were considered "moderately reported," and those reported in >66% of studies were considered "frequently reported"[12].

One-way analysis of variance (ANOVA) was used to evaluate differences in completeness of reporting relative to country, journal, index test type, and subspecialty area. Two-tailed Student's t-test statistics were used to evaluate differences in reporting completeness depending on journal impact factor (median split), use of supplementary material, study design (single test vs. comparative), PRISMA (or extension) citation, and journal's PRISMA adoption status (adopter vs. non-adopter). Correlation of completeness with word count (full text and abstract) was performed by calculating Spearman's rho. These analyses were repeated for PRISMA-DTA for abstracts. The level for statistical significance was set at $P < 0.05$. All statistical analyses were performed using SAS 9.3 (SAS Institute, Cary NC).

Results

Of 881 unique titles and abstracts identified based on our search, 765 were excluded after title and abstract review and 16 after full-text review, resulting in 100 eligible articles included in the current study. The study selection process and reasons for exclusion are outlined in Figure 1. Characteristics of the studies reported in these articles are shown in Table 1. A full list of the included articles along with their PRISMA-DTA and their PRISMA-DTA for abstract completeness is displayed in Appendix 4.

Completeness of reporting relative to PRISMA-DTA

The mean number of PRISMA-DTA items reported was 18.6/26 items (71%, SD=1.9) with a range from 12.0 to 23.0. Figure 2 shows the cumulative completeness of the included articles relative to the number of items.

The completeness of reporting of the 100 study reports on a per-item basis relative to the PRISMA-DTA is summarized in Table 2. Highlights of the detailed Table 2 results as follows: nineteen of the 26 items were frequently reported (>66% of studies) in whole or in part (as sub-items). These include items pertaining to the study selection (Item 9 - methods; Item 17- results), reporting of the statistical methods (Item D2), and the data collection process (Item 10). Twelve of the 26 items were moderately reported (33–66% of studies) in whole or in part, such as items concerned with evaluation of the risk of bias and applicability (Item 19) and the search strategy (Item 8).

Five of the 26 items were infrequently reported (<33% of studies) in whole or in part. These were related to protocol reporting and registration (Item 5), providing definitions used in data extraction (Item 11 - target condition, index test, reference standard), synthesis of the results (Item 14 - methods of handling data, combining results of studies) and characteristics of the included studies (Item 18 - study settings and funding sources). The sum of the number of items reported frequently, moderately, and infrequently was more than 26 as for a given item, some sub-items were present in more than one category.

Completeness of reporting relative to PRISMA-DTA for Abstracts

The mean number of reported items for PRISMA-DTA for abstracts was 5.5/11 items (50%, SD=1.2) with a range from 2.8 to 8.2. Completeness on a per-item basis is summarized in table 3.

Highlights of the detailed table 3 results as follows: five of the 11 items were frequently (>66% of studies) reported in whole or in part. These included items pertaining to the study question (Item 2), number of included studies (Item 6), synthesis and interpretation of results (Items A1, 7 and 10). Seven of the 11 items were moderately reported in whole or in part, such as items concerned with evaluation of the risk of bias and applicability (i.e. Item 5), eligibility criteria (Item 3), and information sources (Item 4). Five of the 11 items were infrequently reported items in whole or in part. These included items relevant to funding information (Item 11), protocol registration (Item 12), strengths and limitations of the Systematic review (Item 9), characteristics of the included studies (Item 6), and assessment of applicability (Item 5).

Subgroup Analysis

A summary of the performed subgroup analyses is presented in Table 4. Variability in reporting by country of the corresponding author was identified (P=0.04); Canadian authors demonstrated the most complete reporting, averaging 20.6/26 items, compared with 17.6/26 items in the country with the lowest number of reported items, China. Completeness of reporting of studies published in higher impact factor journals (median split at 2.768) reported more items than studies published in lower impact factor journals (18.9 vs. 18.1 items, P=0.04). Studies that used supplementary material reported more items than those that did not (19.1 vs. 18.0 items, P=0.004). Studies that cited PRISMA (or extension) reported more items than those that did not

(18.9 vs. 17.7 items, *P*=0.003). No statistical difference in reporting completeness for PRISMA-DTA and PRISMA-DTA for abstracts was identified for the journal of publication (*P*=0.6), limitations of abstract word count by journal (*P*= 0.9), PRISMA adoption by journal (*P*=0.2), structure of abstracts (*P*=0.2), study design (*P*=0.8), subspecialty area (*P*=0.09), or index test (*P*=0.5) (Table 4). Association of completeness with higher word count was present for abstracts (R=0.4; *P*<0.001) but not for full-texts (R= -0.03; *P*=0.06). Additional details on the subgroup analyses performed are presented in Appendix 5.

Discussion

Reports of recent systematic reviews are not fully informative. On average, just over two-thirds of the 26 PRISMA-DTA items were reported in full review reports, with slightly lower proportions for the abstracts of the same articles. Both journal- and study-level variables were associated with completeness: completeness of reporting was higher in journals with higher impact factor and in studies that cited PRISMA (or an extension); however these differences were modest at ~1 item of difference between groups. Limitations imposed by journals may impact completeness of reporting; studies that used supplementary material were more informative, as were abstracts with higher word counts. Variability in reporting was associated with country of corresponding author, with the most reported items observed in studies from Canada and Brazil; however few studies from these countries were identified (<5 each) and the overall difference in items was also modest (<2). China produced double the number of reviews compared to the next most frequent country, and more than a quarter of the systematic reviews included in our analysis; the completeness of reporting of these reviews was the lowest when compared to articles from other countries[22].

Items related to the description of the index test, eligibility criteria, and the study selection process were generally frequently reported. However, items specific to protocol registration, definitions for the data extracted, synthesis of results (methods of handling data and combining results of studies), or the evaluation of risk of bias and applicability for individual studies were infrequently reported[19,23,24]. The lack of transparent reporting in these items limits the ability to assess the validity and generalizability of results[9-11].

Our results show lower completeness of reporting relative to the PRISMA-DTA when compared to evaluations of completeness in imaging systematic reviews (largely DTA) to the

original PRISMA statement examined by Tunis *et al*. These imaging systematic reviews published in radiology journals showed a relatively higher completeness of reporting relative to the PRISMA checklist (81%); this is likely due to the fact that PRISMA-DTA is a new guideline and this is a 'baseline' evaluation rather than a follow up evaluation (Tunis et al. was conducted several years after the publication of PRISMA)[9]. Their analysis identified infrequent reporting in 3 items (Items 5, 15, 22), 2 of which have been omitted from the PRISMA-DTA checklist (Items 15, 22 - examining the risk of bias across studies)[20]. Reporting of the remaining infrequently reported item (Item 5 "protocol and registration") has not improved since 2013: of the 100 included studies, 29 had registered a protocol (all in PROSPERO). This is somewhat perplexing since at the time of Tunis *et al*.'s publication, PROSPERO was relatively new[25]. Clearly, additional measures to encourage protocol registration and reporting are warranted. Conversely, the number of reported PRISMA items was relatively comparable across different subspecialties; Fleming et al. identified completeness of 64% in their recent assessment of systematic reviews in orthodontics[13], Cullis et al. identified completeness of 57% in pediatric surgery systematic reviews[14] , and Gagnier et al. 68% in the orthopedic literature[15].

Two new items were added to the PRISMA-DTA checklist and one was added to the PRISMA-DTA for abstracts checklist [20]: Item D1. "Clinical role of index test", Item D2. "Meta-analysis" (for full-text), and Item A1 "Synthesis of results" (for abstracts). Interestingly, they were all frequently reported. This may be because authors, reviewers, and editors acknowledged the necessity of reporting these items for DTA reviews despite explicit guidance. While the number of reported items was not associated with the length of the publication (word count), the use of supplementary material did influence completeness of reporting. This presents an opportunity: journals not presently offering such a service should consider this potential benefit.

PRISMA citation was associated with more reported items; this is in agreement with a previous study by Page et. al[26]. Citation indicates at least base awareness of the reporting guideline by authors. Corresponding authors of the included studies were contacted and encouraged to use the PRISMA-DTA reporting guidelines in future DTA systematic reviews. Reports in PRISMA adopting journals were not more informative than reports in non-adopting ones. This is discordant with previous studies that have shown that guideline adoption has been associated with better reporting[9,12], and likely related to the fact that 'adoption' was classified based on PRISMA rather than PRISMA-DTA (since PRISMA-DTA was not available at the

time of publication of the reviews). As journal awareness of PRISMA-DTA increases, perhaps the lack of association may change.

This study has several strengths: it included reviews from a wide range of disciplines; a rigorous sub-item scoring system was applied; evaluation for variables associated with completeness was thorough and included many that were not considered in previous assessments of reporting (e.g., word count, supplementary material). However, some potential limitations should be considered. Despite attempts to minimize subjective evaluations of certain items (pilot study, user guide, training meetings), extracted data was inherently subjective and some readers may disagree with the thresholds applied to consider an item 'reported' or not. Furthermore, "not applicable" sub-items were rated as "yes" in the analysis of the collected data. This might have inflated the scores of some "non-adhering" studies; the potential impact is low as this answer ("N/A") was used infrequently and for only a few sub-items (4/68 sub-items).

In conclusion, recently published DTA systematic reviews are not fully informative, when evaluated against the PRISMA-DTA and PRISMA-DTA for abstracts reporting guidelines. Completeness varied by country of the corresponding author and was higher in journals with higher impact factors, studies that cited PRISMA, used supplementary material and had higher abstract word counts. Given that protocol reporting and registration persists as an infrequently reported item, knowledge translation strategies specific to this deficiency should be considered. For example, journals might consider making 'protocol registration' mandatory for systematic reviews as they are for clinical trials in many journals. Journals should also consider practical strategies that would facilitate better reporting; these include providing authors with the opportunity to publish supplementary material and allowing higher word counts in the abstract. These results should guide knowledge translation including journal level (adoption of PRISMA-DTA, increased abstract word count and use of supplementary material), author level (PRISMA-DTA citation-awareness) and other (PRISMA DTA author workshops at Cochrane, PRISMA explanation and elaboration paper) strategies. Follow up evaluations of completeness of reporting can apply the framework and methodology of this study to evaluate changes over time.

Figure Captions

Figure 1. Flow diagram of the included studies

Figure 2. Frequency of reporting of each item of the PRISMA-DTA – 26 items (A) and PRISMA-DTA for abstracts – 11 items (B) for the included articles (n=100). The score for each included article was rounded to the nearest integer.

TABLE 1. Characteristics of Included Articles

Study Characteristics	Number
Country	
China	28
United States of America	14
South Korea	12
United Kingdom	8
Brazil/Canada/Netherlands	4 each
Other	26
Journal	
European Radiology	4
American Journal of Roentgenology	4
BMC infectious diseases	4
Acta Obstetricia et Gynecologica Scandinavica	3
PLOS One	3
The British Journal of Radiology	3
Oncotarget	3
Other	76
Index-test type	
Imaging	58
Laboratory	25
Physical Examination	6
Questionnaire	5
Microbiology	2
Other	4
Subspecialty area	
Diagnostic radiology	40
Laboratory medicine	25
Nuclear Medicine	12
Obstetrics and gynecology	6
Internal Medicine	3
Surgery	2
Microbiology	2
Other	10
Impact Factor	
< 2.768	51
≥ 2.768	49
Study Design	
Single test	65
Comparative	35
Use of Supplementary Material	
No	51
Yes	49
PRISMA citation	
No	30
Yes	70
PRISMA Adoption by journal	
No	64
Yes	36

TABLE 2. Reporting frequency of PRISMA-DTA items. For all included studies **black-shaded items were infrequently reported** (<33%); *gray-shaded items were moderately reported* (33-66% of studies), and unshaded items were frequently reported (>66% of studies)

Item		Sub-Item	Description	Number of studies reporting the item (n=100)
Title	**1**		Identify the report as a systematic review (+/- meta-analysis) of diagnostic test accuracy (DTA) studies	94
Abstract	**2**		Abstract: See PRISMA-DTA for abstracts	
Introduction				
Rationale	**3**		Describe the rationale for the review in the context of what is already known	100
Clinical role of index test	**D1**	D1. a	State the scientific and clinical background, including the intended use and clinical role of the index test	92
		D1. b	If applicable, the rationale for minimally acceptable test accuracy (or minimum difference in accuracy for comparative design) (N/A if no minimal acceptable accuracy specified)	81
Objectives	**4**	4.a	Provide an explicit statement of question(s) being addressed in terms of participants	55
		4.b	Provide an explicit statement of question(s) being addressed in terms of index test (s)	96
		4.c	Provide an explicit statement of question(s) being addressed in terms of target condition(s)	95
Methods				
Protocol and registration	**5**		Indicate where the review protocol can be accessed (e.g., Web address), and, if available, provide registration information including registration number	29

Eligibility criteria	**6**		**Specify study characteristics used as criteria for eligibility, giving rationale for:**	
		6.a	Participants	75
		6.b	Setting	29
		6.c	Index test(s)	96
		6.d	Reference standard(s)	75
		6.e	Target conditions(s)	92
		6.f	Study design	70
		6.g	Report characteristics (e.g., years considered, language, publication status)	88
Information sources	**7**	7.a	Describe all information sources (e.g., contact with study authors to identify additional studies) in the search	87
		7.b	Date last searched	33
Search	**8**		Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated	42
Study selection	**9**		State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)	87
Data collection process	**10**		Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators	84
Definitions for data extraction	**11**		**Provide definitions used in data extraction and classifications of:**	
		11.a	Target condition(s)	21
		11.b	Index test(s)	28
		11.c	Reference standard(s)	18
		11.d	Other characteristics (e.g. study design, clinical setting)	24
Risk of bias and applicability	**12**	12.a	Describe methods used for assessing risk of bias in individual studies	90
		12.b	Describe methods used for assessing concerns regarding the applicability to the review question	71
Diagnostic accuracy measures	**13**	13.a	State the principal diagnostic accuracy measure(s) reported (e.g. sensitivity, specificity)	96
		13.b	State the unit of assessment (e.g. per-patient, per-lesion)	41

Synthesis of results	**14**		**Describe methods of handling data, combining results of studies and describing variability between studies. This could include, but is not limited to:**	
		14.a	Handling of multiple definitions of target condition	26
		14.b	Handling of multiple thresholds of test positivity	37
		14.c	Handling multiple index test readers	31
		14.d	Handling of indeterminate test results	4
		14.e	Grouping and comparing tests	47
		14.f	Handling of different reference standards	22
Meta-analysis	**D2**		Report the statistical methods used for meta-analyses, if performed. (N/A if no meta-analysis done)	90
Additional analyses	**16**	16.a	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done	92
		16.b	Indicate which were pre-specified	43
Results				
Study selection	**17**	17.a	Number of studies screened available	97
		17.b	Number of studies assessed for eligibility available	96
		17.c	Number of studies included in the review available	100
		17.d	Number of studies included in the meta-analysis available, if applicable	100
		17.e	Reasons for exclusions at each stage provided	79
		17.f	Flow diagram provided	94
Study characteristics	**18**		**For each included study provide citations and present key characteristics including:**	
		18.a	Participant characteristics (presentation, prior testing)	67
		18.b	Clinical setting	25
		18.c	Study design	71
		18.d	Target condition definition	45
		18.e	Index test(s)	87
		18.f	Reference standard(s)	62
		18.g	Sample size	94
		18.h	Funding sources	3
Risk of bias and applicability	**19**	19.a	Present evaluation of risk of bias for each study	60

		19.b	Concerns regarding applicability for each study	47
Results of individual studies	**20**		**For each analysis in each study (e.g. unique combination of index test, reference standard, and positivity threshold) report:**	
		20.a	2x2 data (TP, FP, FN, TN)	37
		20.b	Estimates of diagnostic accuracy	83
		20.c	Estimates of confidence intervals	76
		20.d	Forest or ROC plot	87
Synthesis of results	**21**	21.a	Describe test accuracy and meta-analysis results if done	100
		21.b	Describe variability in accuracy (e.g. confidence intervals if meta-analysis done)	100
Additional analyses	**22**		Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression; analysis of index test: failure rates, proportion of inconclusive results, adverse events)	98
Discussion				
Summary	**24**	24.a	Summarize the main findings	98
		24.b	The strength of evidence summarized	54
Limitations	**25**		**Discuss limitations from:**	
		25.a	Included studies (e.g. risk of bias and concerns regarding applicability)	82
		25.b	The review process (e.g. incomplete retrieval of identified research)	51
Conclusions	**26**	26.a	Provide a general interpretation of the results in the context of other evidence	99
		26.b	Discuss implications for future research and clinical practice (e.g. the intended use and clinical role of the index test)	89
Other				
Funding	**27**	27.a	Describe sources of funding for the systematic review and other support	68
		27.b	Describe role of funders for the systematic review (N/A if no funders)	39

TABLE 3. Reporting Frequency of PRISMA-DTA for Abstracts items. For all included studies **black-shaded items were infrequently reported** (<33%); *gray-shaded items were moderately reported* (33-66% of studies), and unshaded items were frequently reported (>66% of studies)

Item		Sub-Item	Description	Number of studies reporting the item (n=100)
Objectives	**2**		The research question including components such as:	
		2.a	Participants	49
		2.b	Index test(s)	99
		2.c	Target condition(s)	97
Methods				
Eligibility criteria	**3**		Study characteristics used as criteria for eligibility	57
Information sources	**4**	4.a	Key databases searched	63
		4.b	Search dates	42
Risk of bias and applicability	**5**	5.a	Methods of assessing risk of bias	38
		5.b	Methods for assessing concerns regarding applicability	25
Synthesis of results	**A1**		Methods for data synthesis	91
Results				
Included studies	**6**	6.a	Number of studies included	96
		6.b	Number of participants included	62
		6.c	Characteristics of included studies (including reference standard)	13
Synthesis of results	**7**		**Results for analysis of diagnostic accuracy:**	
		7.a	Indicate the number of studies	89
		7.b	Indicate the number of participants	62
		7.c	Describe test accuracy (e.g., meta-analysis results if done, if not done, range of accuracies from studies would be a minimum)	88
		7.d	Describe variability (e.g., confidence intervals if meta-analysis was done)	70
Discussion/ Conclusions				
Strengths and Limitations	**9**	9.a	Summary of the strength	8
		9.b	Limitations of the evidence	26
Interpretation	**10**	10.a	General interpretation of the results	96
		10.b	Important implications	58
Other				
Funding	**11**		Primary source of funding for the review.	3
Registration	**12**		Registration number and registry name.	5

TABLE 4. Subgroup analyses evaluating for variability of PRISMA-DTA completeness. Shaded cells are indicative of statistical significance.

Subgroup	Summary of findings	*P*-value
Country	Canada (N= 4; 20.6 items) and Brazil (N= 4; 20.0) reported the most items, while China reported the fewest (N= 28; 17.6)	0.039[1]
Journal	No significant difference in PRISMA-DTA reporting identified	0.584[1]
Index-test type	No significant difference in PRISMA-DTA reporting identified	0.446[1]
Subspecialty area	No significant difference in PRISMA-DTA reporting identified	0.093[1]
Structured abstract	No significant difference in PRISMA-DTA for abstracts reporting identified	0.231[1]
Word limit restriction by journal	No significant difference in PRISMA-DTA for abstracts reporting identified	0.940[1]
Impact Factor	Studies published in higher impact factor journal (relative to the median: 2.768) reported more items than lower impact factor journals (18.9 vs. 18.1 items)	0.038[2]
Study Design	No significant difference in PRISMA-DTA reporting identified	0.785[2]
Use of Supplementary Material	Studies that used supplementary material reported more items than those that did not (19.1 vs. 18.0 items)	0.004[2]
PRISMA citation	Studies that cited PRISMA (or extension) reported more items than those that did not (18.9 vs. 17.7 items)	0.003[2]
Adoption by journal	No significant difference in PRISMA-DTA reporting identified	0.168[2]

1. Analysis of variance (ANOVA) test performed
2. Student's t-test performed

References

1. Balogh E, Miller B, Ball J. *Improving diagnosis in health care.* Institute of Medicine (U.S.). Committee on Diagnostic Error in Health Care.: Washington, DC: The National Academies Press; 2015.
2. Singh H, Graber ML. Improving Diagnosis in Health Care--The Next Imperative for Patient Safety. *N Engl J Med.* 2015;373(26):2493-2495.
3. 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.
4. deVet HCWEA, Riphagen II, Aertgeerts B, Pewsner D. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 0.4; 2008. http://www.cochrane.org/editorial-and-publishing-policy-resource/cochrane-handbook-diagnostic-test-accuracy-reviews
5. 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.
6. 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.
7. 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.
8. McInnes MD, Bossuyt PM. Pitfalls of Systematic Reviews and Meta-Analyses in Imaging Research. *Radiology.* 2015;277(1):13-21.
9. 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.
10. 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.
11. 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.
12. Hong PJ, Korevaar DA, McGrath TA, et al. Reporting of imaging diagnostic accuracy studies with focus on MRI subgroup: Adherence to STARD 2015. *J Magn Reson Imaging.* 2017.
13. Fleming PS, Seehra J, Polychronopoulou A, Fedorowicz Z, Pandis N. A PRISMA assessment of the reporting quality of systematic reviews in orthodontics. *Angle Orthod.* 2013;83(1):158-163.
14. Cullis PS, Gudlaugsdottir K, Andrews J. A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PLoS One.* 2017;12(4):e0175213.
15. Gagnier JJ, Mullins M, Huang H, et al. A Systematic Review of Measurement Properties of Patient-Reported Outcome Measures Used in Patients Undergoing Total Knee Arthroplasty. *J Arthroplasty.* 2017;32(5):1688-1697.e1687.

16. Kelly SE, Moher D, Clifford TJ. Quality of conduct and reporting in rapid reviews: an exploration of compliance with PRISMA and AMSTAR guidelines. *Syst Rev.* 2016;5:79.
17. Equator Network. Reporting guidelines under development. http://www.equator-network.org. Accessed may 13, 2018.
18. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.* 2009;339:b2535.
19. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155(8):529-536.
20. McInnes MDF, Moher D, Thombs BD, et al. Preferred Reporting Items for a Systematic Review and Meta-analysis of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement. *JAMA.* 2018;319(4):388-396.
21. Korevaar DA, Cohen JF, Reitsma JB, et al. Updating standards for reporting diagnostic accuracy: the development of STARD 2015. *Res Integr Peer Rev.* 2016;1:7.
22. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *Milbank Q.* 2016;94(3):485-514.
23. Moher D, Tetzlaff J, Tricco AC, Sampson M, Altman DG. Epidemiology and reporting characteristics of systematic reviews. *PLoS Med.* 2007;4(3):e78.
24. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Group Q-S. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol.* 2013;66(10):1093-1104.
25. PROSPERO - International prospective register of systematic reviews. NHS National Institute for Health Research. http://www.crd.york.ac.uk/prospero/. Accessed 13 May 2018.
26. Page MJ, Shamseer L, Altman DG, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS Med.* 2016;13(5):e1002028.