ELSEVIER

Seminars in
NUCLEAR
MEDICINE

Reporting Guidelines for Imaging Research

Matthew D.F. McInnes, MD, PhD,* Christopher S. Lim, BMBS, FRCPC,† Christian B. van der Pol, MD, FRCPC,‡ Jean-Paul Salameh, BSc,§ Trevor A. McGrath, MD,‖ and Robert A. Frank, MD‖

Suboptimal reporting in the publication of imaging research studies is a growing concern. Deficient and incomplete reporting prevents the evaluation of the validity, replicability, and the overall quality of the research. Reporting guidelines are checklists designed to guide researchers about the minimum information to be provided in research studies to allow for adequate quality appraisal and to assess generalizability. They are a powerful tool to allow key stakeholders such as journal editors, peer reviewers, funding agencies, and readers to better identify robust health research. The Enhancing the QUAlity and Transparency of Health Research Network is an international initiative that attempts to improve the reporting practices of a variety of health research study designs by providing the resources required to develop, disseminate, and implement reporting guidelines. In this review, we elaborate on the impact of good reporting on imaging research, and the different types of guidelines relevant for the various study designs applicable in imaging research.
Semin Nucl Med 49:121-135

Introduction

Efficient implementation of the latest research findings into clinical imaging practice relies on availability of high-level evidence. Incomplete reporting of health research has raised concerns regarding the ability to evaluate validity of research needed to appropriately apply to clinical practice. Insufficient reporting has been identified in several areas, including selective outcome reporting,[1,2] misinterpretation, and misrepresentation of data,[3,4] inadequate reporting of harms,[5] and insufficient methodological description.[6,7] These deficiencies prevent stakeholders who rely on health-research from critically assessing the quality of evidence and may also promote misallocation of resources.[8,9] In nuclear medicine, the impact may be felt through inappropriate implementation of diagnostic tests based on biased evidence and may contribute to harm from false negative (missed disease) or false positive (over-diagnosis) test results.

Reporting guidelines are checklists designed to guide researchers about the minimum information to be provided in research studies in order to ensure there is sufficient detail to allow for adequate quality appraisal and to assess generalizability. These guidelines have the potential to improve transparency and completeness of reporting—prerequisites to reproducibility.[10-13] Several studies evaluating the impact of adherence to reporting guidelines on the completeness of reporting, and consequently on the quality of research have shown improvements in reporting of randomized control trials (RCTs),[14] diagnostic test accuracy (DTA) studies,[15,16] and systematic reviews.[7]

The Enhancing the QUAlity and Transparency of Health Research (EQUATOR) Network is an initiative that seeks to improve the quality of reporting by promoting complete and transparent reporting of health research studies. The network is a global program that provides training, resources, and assistance for the development, dissemination, and implementation of reporting guidelines in all areas of health research.[17] The EQUATOR library (http://www.equator-network.org/library) provides a searchable database of all published reporting guidelines (since 1996).[18]

With the growing role of imaging research, including nuclear medicine, in evidence-based medicine, high-quality

*Rm c-159 Department of Medical Imaging, The Ottawa Hospital—Civic Campus, Ottawa, ON, Canada.
†Department of Medical Imaging, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, Canada.
‡Juravinski Hospital and Cancer Centre, Hamilton Health Sciences, McMaster University, Hamilton, ON, Canada.
§The University of Ottawa, School of Epidemiology and Public Health, The Ottawa Hospital Research Institute Clinical Epidemiology Program, Ottawa, Canada.
‖The University of Ottawa, Department of Radiology, Ottawa, Canada.
Address reprint requests to Matthew D.F. McInnes, MD, PhD, Rm c-159 Department of Medical Imaging, The Ottawa Hospital—Civic Campus, 1053 Carling Ave, Ottawa, ON K1E 4Y9. E-mail: mmcinnes@toh.ca

https://doi.org/10.1053/j.semnuclmed.2018.11.008
0001-2998/

reporting of nuclear medicine studies is not an optional extra, but an essential element (Table 1). Imaging research covers a wide spectrum of study designs ranging from RCTs, to DTA studies, and systematic reviews. The unique methodological approaches for each study design entail specific reporting requirements. In this article, we discuss the different reporting guidelines relevant to the different types of studies in imaging research: Standards for Reporting of Diagnostic Accuracy (STARD) 2015 for DTA studies,[13] preferred reporting items for systematic reviews and meta-analyses (PRISMA) for systematic reviews[10] and PRISMA-DTA for DTA systematic reviews[11]; and the Consolidated Standards of Reporting Trials (CONSORT) statement for RCTs[12] (Table 2 – Appendix 1).

STARD 2015 for DTA Studies

Diagnostic test accuracy is a common study design in nuclear medicine research.[16,19-21] Studies of DTA compare outcomes from one or more index tests to those obtained from a reference standard. Cross-classification of the index test results against the reference standard (true positive, false negative, true negative, false positive) allows investigators to express diagnostic accuracy estimates in various ways: sensitivity and specificity; positive and negative predictive values; likelihood ratios; area under the receiver operator characteristic curve; and diagnostic odds ratio. Knowledge of the characteristics for a nuclear medicine imaging test is important for guiding clinical decision-making. For instance, Lodge et al. found 100% sensitivity and 76% specificity for 18-FDG PET in discriminating malignant from benign soft tissue masses.[22] These findings could rule out malignancy with a negative result, but a positive result might require confirmatory follow-up testing.[22] However, the validity of these estimates may be compromised by several methodological factors (eg participant recruitment, data collection, and execution or interpretation of tests or data analysis), which must be taken into consideration when interpreting a diagnostic accuracy test result.[23]

Table 1 Benefits of Reporting Guidelines for Research

Stakeholders	Benefits
Journal editors	Facilitate the review process by allowing easy evaluation of the methodological quality of research and validity of the findings
Funding agencies	Promote fair and adequate allocation of resources by research funders
	Improve transparency, and completeness of reporting
Researchers	Guide researchers through the crafting stages of the study protocol
	Improve the quality of research
	Facilitate efforts for reproducibility
	Increase ability to assess generalizability of study findings

Table 2 Reporting Guidelines Relevant to Nuclear Medicine Research

Type of Study	Guideline Name/Acronym
Randomized controlled trials (RCTs)	CONSORT Statement
	CONSORT Harms: Reporting harms in RCTs
	CONSORT Non-inferiority: Reporting non- inferiority and equivalence RCTs
Dosimetry Reporting	EANM Dosimetry Committee guidance document: good practice of clinical dosimetry reporting
Diagnostic test accuracy (DTA) studies	STARD 2015
Systematic reviews and meta-analyses	PRISMA
	PRISMA of Diagnostic Test Accuracy Studies: The PRISMA-DTA Statement
	PRISMA harms checklist: improving harms reporting in systematic reviews
	PRISMA of individual participant data: the PRISMA-IPD Statement

Shaded cells represent extensions of the original statements. Full statements can be found on the EQUATOR website (http://www.equator-network.org).

The STARD statement is a checklist of the minimum essential items to be reported in all studies assessing the accuracy of a diagnostic test. The original STARD statement was developed in 2003 with the aim of improving the transparency and completeness of diagnostic accuracy study reporting, thereby enabling readers to assess generalizability, applicability, validity, and potentials for bias.[24] In 2015, the STARD committee published an updated checklist (STARD 2015) incorporating updated evidence, optimizing the utility of the checklist, and harmonizing the checklist with other reporting guidelines such as CONSORT.[13] STARD 2015 was complemented with an explanation and elaboration document clarifying the rationale behind each item and facilitating proper interpretation and use of the checklist.[25] The 30-item STARD 2015 checklist is open-access and available from the EQUATOR website (http://www.equator-network.org/library).

The STARD initiative has been well-received since its initial dissemination, widely endorsed by journals and cited by authors.[26] The original checklist was published simultaneously in 24 journals,[24] and the subsequent 2015 update was published in three.[13,27,28] Together, the STARD checklists have been endorsed by more than 120 journals[29] (including at least 15 radiology, nuclear medicine, and medical imaging journals as defined by Thomson-Reuters[30,31]) and cited in over 2500 studies.[32]

There are several items on the STARD 2015 checklist which are particularly relevant to nuclear medicine DTA

research (ie items 3, 10, 12, 13, 22, and 25). One such item advises authors to clearly outline the intended use and clinical role of the index test (item 3).[13] It is important that readers are aware of whether a study is evaluating the accuracy of an imaging test for diagnosis, screening, staging, surveillance, or prognosis in order to determine the appropriateness of the methodology as well as the utility of the index test in specific clinical scenarios. Specifying the anticipated placement of the test within the established clinical pathway — an initial triage test, adjunctive test, or replacement of a current standard — also has important implications on the interpretation of the accuracy results. For example, bone scintigraphy is useful in the screening and detection of bone metastasis due to its high sensitivity, full skeletal coverage, and widespread availability[33-35]; allowing for a trade-off with a lower specificity.[36] However, if the test is further along in the clinical pathway and is to be used to rule in disease, a high specificity will be required, such as MRI.[37]

Another important item of STARD is reporting of the index test and reference standard in sufficient detail to allow replication, including the expertise, training level, and number of individuals assessing the index and reference tests (item 10).[13] Compared to quantitative laboratory tests, diagnostic imaging tests have an additional element of subjectivity owing to their qualitative nature. Image interpretation can vary depending on the expertise and experience of the assessor and inter-reader variability exists.[38,39] Knowledge of the training level of the individuals assessing both the index and reference tests is essential in allowing readers to appraise the results of a study and to judge whether the results are applicable to their clinical settings.

Readers should also be aware of whether the individuals interpreting the index test and reference standard were blinded to other clinical information (item 13).[13] Again, while this is less relevant for quantitative tests, knowledge of the patient's clinical information or other imaging results can confer substantial bias.[40] In standard diagnostic radiology and nuclear medicine practice, imaging studies are interpreted in the context of clinical history and previous imaging, when available. It is particularly important in diagnostic accuracy studies to report whether the study was performed in a controlled setting (ie, restricting the available information in an experimental setting) or whether images were assessed as part of routine clinical practice (ie, without blinding), as this can have a major impact on the generalizability of the final result.

Defining test positivity and result classifications is another key area of reporting in studies of imaging diagnostic test accuracy (item 12).[13] The potential for subjectivity in imaging research makes it critical for authors to provide clear definitions of imaging findings associated with positive or negative results — this information is necessary for reproduction and clinical application of the study results. For example, studies assessing the diagnostic accuracy of 18-FDG-PET/CT in the setting of solitary pulmonary nodules and the evaluation of lung cancer, may assign different SUV_{max} positive cut points when classifying results[41]; as such, it is crucial that readers are aware of these differences in order to accurately interpret between-study differences. A critical element in the reporting of this item is to clarify whether positivity cut-offs or result categories were determined *a priori* (ie, pre-specified) or *post-hoc* (by exploring the data to optimize thresholds), as the latter approach leads to overestimated diagnostic accuracy and results that are often not replicated.[42,43]

Authors are advised to report the time interval and any clinical interventions between the index test and reference standard (item 22).[13] Although an ideal diagnostic accuracy study-design is cross-sectional (ie, index and reference tests performed in the same patients, at the same time),[44] these conditions are often not feasible. Delays between the index test and reference standard may increase the likelihood of alterations in the disease process due to the natural history of disease or there may be intervening treatment, which can confound the observed results. Such delays are common in nuclear medicine imaging research, especially when the reference standard is a follow-up imaging test. Cancer can be dynamic in that a patient's disease status can change substantially over time due to tumor progression or response to treatment. Accordingly, it is of particular importance in nuclear medicine research to specify the time interval (and treatment regimen) occurring between the index and reference tests, as any intervening time can substantially bias the estimated accuracy.

Lastly, an important consideration in diagnostic accuracy research is the occurrence of adverse reactions (item 25).[13] While imaging is considered to be a relatively safe intervention, the potential for adverse events exists and remains an important consideration in balancing harms vs benefits of a test. While reported adverse reactions to diagnostic radiopharmaceuticals are uncommon and predominantly minor,[45] these events might be underreported and severe adverse events can occur, including anaphylactoid reactions and death.[46]

Publication of the original STARD checklist was met with modest improvements in the reporting of diagnostic accuracy studies.[26] A recent study assessing adherence of imaging diagnostic accuracy studies to STARD 2015 found moderate adherence overall and identified multiple key areas of deficiency.[16] Several of the items that were infrequently reported with particular relevance to imaging include: adverse events, timing between index test and reference standard, blinding to clinical information and previous testing, and outcome categorization. Among the least frequently reported items was the adverse effects of the index test and reference standard, which were reported in <5% of studies.[16] There was moderate reporting of the time interval between the index and reference tests (39%), but few studies reported the occurrence of intervening clinical interventions (15%).[16] A moderate proportion of studies reported reader training level (60%) as well as blinding of index test readers to the reference standard and clinical information (39%-40%), which was higher than reporting rates for blinding of reference test readers to the index test (15%) and clinical data (9%).[16] While the majority of studies (86%) reported positivity cut-offs for the index test, only 36% reported whether these were

determined *a priori* or *post-hoc*, leaving a major barrier to the assessment of bias in the results of most studies.[16] Overall, nuclear medicine was the most highly represented subspecialty in this analysis 32/142 (23%), with a mean of 16.3/30 (54%) reported items among these studies. A prior dedicated assessment of nuclear medicine study adherence to the 2003 STARD checklist showed slightly superior reporting 17/25 items (68%).[21] These findings can guide nuclear medicine DTA authors, editors, reviewers and journals on areas of deficiency to focus on to improve reporting of future research.

STARD 2015 is not specific to nuclear medicine and imaging studies; rather, it was developed as a general guideline applicable to all areas of diagnostic accuracy research, including laboratory medicine, physical exams, and screening questionnaires. Although this approach was taken in order to maximize dissemination and acceptance by authors, reviewers, and journal editors, it also necessitated that the items remain broadly applicable to all diagnostic tests.[13] Thus, a limitation of STARD 2015 is that it does not provide targeted guidance specific to certain tests or study designs, and therefore must function as a minimum reporting recommendation, rather than a comprehensive guideline. There are several unique imaging-related details that are not specifically mentioned in the general STARD 2015 checklist and elaboration document. For example, when describing details of the index test, proper reporting should include modality, vendor, technical parameters used, type and dose of contrast used, radiopharmaceutical administration protocol, concentration and dose, and the method of handling multiple readers (eg consensus vs independent and measures of interobserver agreement, if applicable). Accordingly, the STARD group encourages the development of checklist extensions specific to certain topics.[13] Several extensions are available on the EQUATOR website,[47-50] but there is not yet a STARD extension specific to imaging or nuclear medicine diagnostic test accuracy studies. An imaging extension of STARD may improve the completeness of reporting specific to nuclear medicine.

PRISMA and PRISMA-DTA for Systematic Reviews

The Cochrane Collaboration defines a systematic review as a review of published evidence with a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included within the review.[51] The completeness of reporting of systematic reviews in imaging research has been identified as being suboptimal.[7,52,53] The PRISMA statement was published in 2009 with the aim of improving reporting of systematic reviews and reducing research waste.[10] Encouragingly, there has been an overall improvement in the completeness of reporting in published systematic reviews since the publication of the PRISMA statement.[54]

The most common type of systematic review in imaging research is DTA[55]; one hundred twenty-four nuclear medicine systematic reviews have been published between 2000-2016 and the rate of publication for these reviews is trending upward.[55] Despite the improvements in overall completeness of reporting in systematic reviews, there are several methodological differences and thus differences in required reporting for DTA systematic reviews compared to systematic reviews of interventions such as pharmaceuticals.[3,19,54,56] While the PRISMA statement of Moher et al. was a step forward for completeness of reporting for all systematic reviews, DTA-specific systematic review reporting guidance was still lacking.[53,57] In 2018, a 27-item extension of the PRISMA statement for diagnostic test accuracy (PRISMA-DTA) was published along with a companion 12-item guideline for abstracts of DTA systematic reviews.[11] To reflect DTA-specific or contemporary best practices for systematic review reporting, 8 of the 27 PRISMA items were unaltered, 17 were modified, 2 were added, and 2 were removed.[11] This new 27-item DTA-specific reporting guideline should be used for DTA systematic reviews rather than the original 2009 PRISMA statement.

Many of the modifications to the original PRISMA items were done to tailor them to DTA systematic reviews. The familiar participants, intervention, comparison, outcomes terminology was replaced with the more apt participants, index test, target conditions (PIT). Data analysis decisions such as handling multiple readers of an index test or multiple thresholds for positivity of an index test are specific to DTA systematic reviews and are new to the PRISMA-DTA guideline. Reporting accuracy estimates with confidence intervals for each included primary study along with 2 × 2 true positive, false negative, false positive, true negative data is also new to PRISMA-DTA. These modifications are crucial to clear interpretation and reproducibility of DTA systematic reviews.

The first of two new items added to PRISMA-DTA is to state the intended use and clinical role of the index test (Item D1).[11] This reporting aspect was found to be deficient in nearly half of recent DTA systematic reviews that were assessed.[3] Also relevant to the STARD statement and discussed above, diagnostic testing can be used for many purposes such as screening, risk stratification, staging, prognosis, treatment selection, or surveillance. The clinical role of the index test describes its position relative to existing tests for the same purpose in the same clinical setting: triage (test used for initial assessment), add-on (test added to the traditional pathway), or a replacement for an existing test or test strategy.[25,58] The intended use and clinical role of the index test will determine the requisite levels of diagnostic accuracy. Clearly stating the intended use and clinical role of the index test will allow readers of DTA systematic reviews to judge the value of the review findings within their own clinical scenario.

The second of the new items added to PRISMA-DTA is to state the statistical methods used for pooling results across studies (meta-analysis) if performed.[11] In many cases, a DTA systematic review will contain a meta-analysis to provide

summary estimates of accuracy measures, often sensitivity and specificity, along with associated confidence intervals for the index test being reviewed. Since sensitivity and specificity are correlated, hierarchical statistical methods, such as the bivariate random effects model or the hierarchical summary receiver operating characteristic model, are currently recommended.[59-61] It has been shown that using univariate meta-analysis methods (ie without considering correlation) rather than the recommended hierarchical methods results in overestimation of summary estimates and narrower confidence intervals.[19] For example, a review assessing the ability of 18F-FDG PET to detect unknown primary tumors reported a summary sensitivity of 0.78 (95% CI 0.72-0.84) using univariate meta-analysis methods.[62] When the data for this study was re-analyzed using a bivariate random-effects (hierarchical) model the summary sensitivity was 0.70 (95% CI 0.65-0.76).[19] Due to this, it is important to report the methods used for meta-analysis within a review, so the reader may assess any bias introduced by nonrecommended statistical methods.

The PRISMA-DTA guideline is general to all DTA systematic reviews and therefore, similar to the STARD guidelines, has limitations when applied to imaging research. The items listed in PRISMA-DTA are the minimally acceptable reporting standards for transparency and reproducibility. There are several items specific to imaging, which are not explicitly mentioned in PRISMA-DTA and also not addressed in STARD. Details such as the modality, vendor, imaging protocol (slice thickness, tube voltage, magnetic field strength, MR sequences used) and radiopharmaceutical concentration and dose are essential information for imaging systematic reviews, but not applicable to all DTA reviews, and thus are not required by PRISMA-DTA.

There are unique challenges associated with DTA systematic reviews not covered by PRISMA and therefore the PRISMA-DTA extension was published.[11] This is currently the best practice reporting guideline, which should be used for imaging systematic reviews. Both PRISMA-DTA and PRISMA-DTA for abstracts are freely available for download on the EQUATOR website.[18]

CONSORT for Clinical Trials Reporting

The CONSORT statement is an evidence-based itemized checklist and flow diagram for reporting of randomized trials[63,64] (Appendix 1). Well-designed, conducted and reported randomized trials represent the optimal design for evaluation of health care interventions.[63,65] However, if a randomized trial lacks logical planning, is not meticulously conducted or is improperly analyzed, the study can be prone to error leading to biased results.[63,65] Assessing the quality of a randomized study is multidimensional, with numerous scales available making logical comparison difficult.[66] The development of CONSORT was created to offer a standardized way for authors to report trial findings, ensure complete and transparent reporting and allow readers to critically appraise each individual randomized trial.[67] For example, the use of 18F-Flucivlocine PET/CT is increasingly being explored in patients with biochemically recurrent prostate cancer.[68,69] Jani et al. found, in their randomized trial which strives to adhere to CONSORT, that information from 18F-Flucivlocine PET/CT led to differences in the radiation treatment volumes and increased radiation dose to the penile bulb in patients undergoing salvage radiotherapy for prostate cancer.[70] Likewise, the randomized trial by Lebech et al. found higher specificity and accuracy of 18F-FDG PET/CT, compared to conventional CT, for detecting cancer in patients with serious symptoms concerning for cancer.[71] Well conducted randomized trials continue to be published in leading nuclear medicine journals encouraging and endorsing the use of CONSORT.[67]

Intuitively, the measure of the methodological quality of a randomized trial is related to the quality of reporting.[66] The CONSORT 2010 statement and the explanation and elaboration document provide guidance for reporting parallel group randomized trials.[63,64,67] The subsequently developed CONSORT extensions provide reporting guidance on most other types of randomized trials.[63,64,67] The CONSORT checklist items facilitate clarity, completeness, and transparency to decrease omissions and ambiguity in the study reporting.[63,64] Although the goal of the CONSORT statement does not include recommendations for study methodology, conduct or analysis, the CONSORT 2010 explanation and elaboration document touches up on the study design, conduct, and publication.[63,64] This had led to numerous journals, conferences, and editorial groups, including *the Journal of Nuclear Medicine*, endorsing and encouraging the use of CONSORT in the reporting of randomized trials.[67,72,73] CONSORT is constantly evolving and modifying items based on the current evidence available.[63,64] The CONSORT 25 itemized checklist is divided into 6 sections describing what information is necessary to include when reporting RCTs.[18,63] The accompanying explanation and elaboration document expands on the specific rationale and justifications of the 25 checklist points with specific examples to help investigators write or appraise trial reports.[64] Additionally, the CONSORT checklist also contains a patient flow diagram which allows clear and transparent graphical information to be displayed about how the trial was conducted, report enrollment, follow-up and analysis of patients involved, which would otherwise be difficult to convey solely using text.[18,63]

Clear and concise information is important and recommended by CONSORT when reporting the title and abstract. These sections are important as some readers only assess an article based on the title and abstract while other readers screen these sections to decide if the article is relevant to read and appraise. As such, the title and abstract should be clear, transparent, and have sufficient detail to serve as an accurate record of its conduct and findings.

The methods section should address 10 subsections described in the CONSORT checklist, and is important because it is a major determinant of the internal validity of a

study. Despite its importance, methods describing participant assignment has been shown to be repeatedly deficient and was only reported in 34% of 616 reports indexed in 2006.[64,74] Proper randomization (eg using a computerized random number generator) and allocation concealment from the researchers assigned to patient enrollment is essential to reducing selection bias in randomized trials.[64] As such, the details of any randomized restriction (eg block size) should be explained and justified.

One of the most important methodological subsections is *the eligibility criteria for participants*. This subsection is essential in understanding the study population and to whom the results of the study apply. A thorough description of the eligibility inclusion and exclusion criteria is important for the readers to understand the context and in the interpretation of the study. CONSORT also strongly recommends the use of a flow diagram as participation flow can be complex and difficult to describe, particularly when patients are lost to follow up, excluded from analysis or do not receive the allocated treatment. In studies that included a flow diagram, more complete reporting of randomized trials was found supporting its recommendations in the CONSORT statement.[75] The use of the flow diagram can also help convey and explain any deviations from the intended protocol.

CONSORT publications have been cumulatively cited over 8000 times and have been listed among the top health research milestones of the 20th century according to the Patient-Centered Outcomes Research Institute.[67,76,77] Since the implementation of CONSORT, studies have demonstrated improved quality of reporting of randomized studies.[14,78-81] Although reporting may be more complete since the development and widespread use of CONSORT, as discussed above, several facets of the CONSORT guidelines remain poorly reported.[67,80,81] The enhanced reporting of randomized trials since CONSORT implementation is encouraging and may improve the quality of randomized studies; however, there remains significant room for improvement.

The EQUATOR networks helps facilitate CONSORT reporting guideline development and dissemination.[63] Apart from CONSORT, there are a number of other guidelines relevant to imaging and nuclear medicine endorsed and available on the EQUATOR website including the EANM dosimetry committee guidance document: good practice of clinical dosimetry reporting.[82]

Conclusion and Future Directions

Reporting guidelines help direct authors to provide clarity on study methodology, which is useful for readers to understand the scope of a study, potential sources of bias, and the generalizability of study findings with respect to their own practice. This includes medical imaging physicians, specifically nuclear medicine specialists and radiologists, who work with a wide variety of patient populations in many different clinical settings and with virtually all medical specialties.

Guidelines for certain study types (such as observational studies, case reports...) may not have been specifically addressed in this article. A list of study types and relevant reporting guidelines can be found on the EQUATOR network website. Several current reporting guidelines for imaging diagnostic accuracy studies were reviewed, including STARD 2015[13] and PRISMA-DTA[11] as well as the CONSORT guidelines for reporting of RCT[12]; it should be noted that these are not specific to nuclear medicine, and should be considered the minimum that is needed to be reported. Additional details relevant to nuclear medicine diagnostic accuracy research which are not specifically discussed in these guidelines include (but are not limited to) modality, vendor, imaging protocol (including specific PET, SPECT, and planar imaging parameters including time, count, collimation and windowing; CT parameters were relevant such as tube current, tube voltage, slice thickness; and specific MRI parameters were relevant such as magnetic field strength, MR sequences, slice thickness), radiopharmaceutical dosage, and the measures of interobserver agreement when multiple readers were used.[56] Stating the study "level" of analysis is also helpful. This is a common issue for imaging studies assessing diseases with a propensity toward multifocality. Consider the example of assessment for metastatic lymph nodes from pelvic malignancies. It is difficult, if not impossible, to directly match individual nodes on imaging to their correlates on surgical histopathology specimens, which can contain tens or evens hundreds of resected nodes. Studies assessing the accuracy of imaging in this setting are typically on a per-patient, per-pelvic side, or per-nodal station basis; evaluations made on per-lesion rather than per-patient basis require specific methods for analysis (typically referred to as "mixed methods models") to account for clustering effects.[83] Several of these points are raised in the more recent guideline for diagnostic test accuracy studies described earlier, the PRISMA-DTA statement, which applies to systematic reviews and meta-analyses.[11] The Cochrane Diagnostic Test Accuracy Working Group are producing a handbook for guidance on reviews of diagnostic test accuracy studies, which may provide insights on methodology applicable not only for reviews but also for individual diagnostic test accuracy studies.[84]

Several developments may be of relevance to future reporting guidelines. The application of machine learning and artificial intelligence in the interpretation of imaging exams is one such example; reporting guidelines would be useful for readers to assess the quality and applicability of algorithms to their work. Several algorithms such as deep learning with convolutional neural networks function as a "black box" in the sense that their inner workings are not fully understood, even if the outputs are predictable. For instance, in the case of observed differences in the detection rates of two similar algorithms for the detection of lung nodules, no concrete guidance is available yet for the assessment of the real reason for this difference; while one algorithm could truly be better, it could be argued that one is only detecting somewhat easier

cases. Guidelines clarifying important parameters to report for these algorithms may be useful; an extension of STARD for this type of research would be helpful to ensure that such relevant details are reported. For systematic reviews and meta-analyses, this could include an extension to the PRISMA statement. Similar extensions exist for individual patient data and network meta-analysis, which are becoming increasingly popular.[85-87]

Unfortunately, adherence to reporting guidelines is suboptimal and room for improvement remains.[16,87,88] Adoption and endorsement of reporting guidelines by journals may improve adherence by authors.[81,89] This has precedent for other items deemed important to a journal's readership; for example, many journals require authors to state conflicts of interest, sources of research funding, and for those conducting trials on human or animal subjects, that research was approved by a local research ethics boards. Investigators hoping to submit manuscripts to these journals know they must first receive research ethics board approval prior to beginning a study, otherwise manuscripts can be rejected or revoked. Other approaches to improve reporting guideline adherence would be to shift the responsibility of completing a checklist to article reviewers rather than authors, who could then request missing information from the study authors. This would however place more demands on those with less "vested interest" in a study, and the extra time commitment could conceivably deter potential reviewers. Journal editorial offices could also assume responsibility for completing a checklist on reporting guideline adherence, however again this would place increased demands on editorial offices. A central registry of reporting guideline adherence checklists may be useful.

Study reporting may improve as reporting guidelines become more ubiquitous. Tunis et al. found that, following publication of the PRISMA guidelines, there was improved completeness of reporting of systematic reviews and meta-analyses in high impact factor radiology journals.[7] Evidence on the correlation between adherence to reporting guidelines and study impact is more conflicting. van der Pol et al. found that there was a positive correlation between both the quality and completeness of reporting (PRISMA) of systematic reviews and meta-analyses and citation rate.[90] However, Dilauro et al. found no association between adherence to the STARD statement and article citation rate.[91] Further research clarifying the utility of reporting guidelines would be beneficial.

Quality scientific inquiry demands an objective approach and sound methodology, and is inextricably linked to clear and comprehensive reporting. Reporting guidelines provide a roadmap to all involved with the publication process. Guidelines can help investigators decide on methodology at study onset. Moving forward, it is our hope that reporting guidelines for studies related to nuclear medicine will become more broadly adopted by journals, reviewers, and most importantly investigators.

Appendix 1. PRISMA Checklist

Section/Topic	#	Checklist Item
TITLE		
Title	1	Identify the report as a systematic review, meta-analysis, or both.
ABSTRACT		
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.
INTRODUCTION		
Rationale	3	Describe the rationale for the review in the context of what is already known.
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (eg, Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (eg, PICOS, length of follow-up) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.
Study selection	9	State the process for selecting studies (ie, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).

(*Continued*)

Section/Topic	#	Checklist Item
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Data items	11	List and define all variables for which data were sought (eg, PICOS, funding sources) and any assumptions and simplifications made.
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.
Summary measures	13	State the principal summary measures (eg, risk ratio, difference in means).
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (eg, I^2) for each meta-analysis.
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (eg, publication bias, selective reporting within studies).
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.
RESULTS		
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each study, present characteristics for which data were extracted (eg, study size, PICOS, follow-up period) and provide the citations.
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).
Additional analysis	23	Give results of additional analyses, if done (eg, sensitivity or subgroup analyses, meta-regression [see Item 16]).
DISCUSSION		
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (eg, healthcare providers, users, and policy makers).
Limitations	25	Discuss limitations at study and outcome level (eg, risk of bias), and at review-level (eg, incomplete retrieval of identified research, reporting bias).
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.
FUNDING		
Funding	27	Describe sources of funding for the systematic review and other support (eg, supply of data); role of funders for the systematic review.

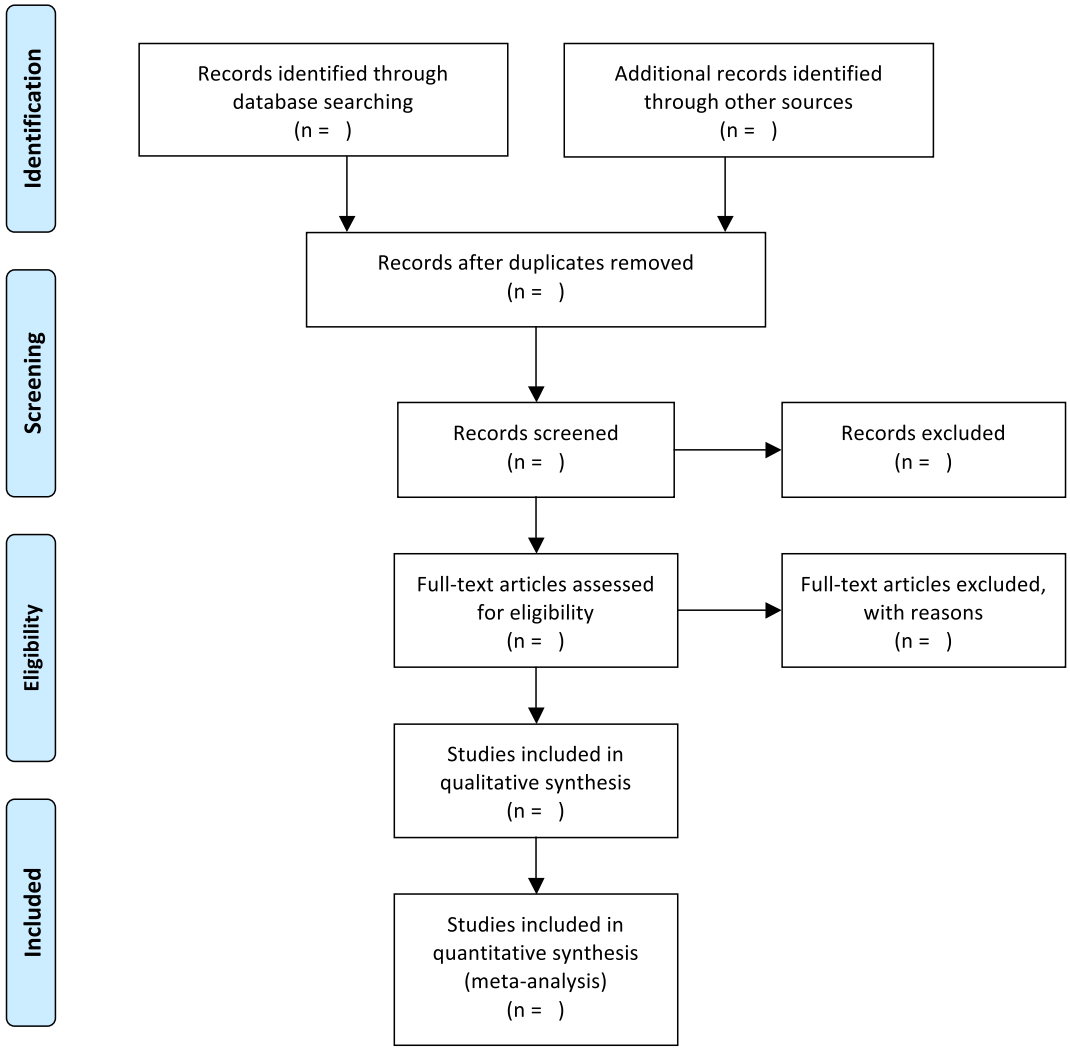
PRISMA-DTA Checklist

Section/topic	#	PRISMA-DTA Checklist Item
TITLE/ABSTRACT		
Title	1	Identify the report as a systematic review (+/− meta-analysis) of diagnostic test accuracy (DTA) studies.
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.
Clinical role of index test	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 comparative design).
Objectives	4	Provide an explicit statement of question(s) being addressed in terms of participants, index test(s), and target condition(s).

(*Continued*)

Section/topic	#	PRISMA-DTA Checklist Item
METHODS		
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (eg, Web address), and, if available, provide registration information including registration number.
Eligibility criteria	6	Specify study characteristics (participants, setting, index test(s), reference standard(s), target condition(s), and study design) and report characteristics (eg, years considered, language, publication status) used as criteria for eligibility, giving rationale.
Information sources	7	Describe all information sources (eg, databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.
Search	8	Present full search strategies for all electronic databases and other sources searched, including any limits used, such that they could be repeated.
Study selection	9	State the process for selecting studies (ie, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).
Data collection process	10	Describe method of data extraction from reports (eg, piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.
Definitions for data extraction	11	Provide definitions used in data extraction and classifications of target condition(s), index test(s), reference standard(s) and other characteristics (eg study design, clinical setting).
Risk of bias and applicability	12	Describe methods used for assessing risk of bias in individual studies and concerns regarding the applicability to the review question.
Diagnostic accuracy measures	13	State the principal diagnostic accuracy measure(s) reported (eg sensitivity, specificity) and state the unit of assessment (eg per-patient, per-lesion).
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: (a) handling of multiple definitions of target condition. (b) handling of multiple thresholds of test positivity, (c) handling multiple index test readers, (d) handling of indeterminate test results, (e) grouping and comparing tests, (f) handling of different reference standards
Meta-analysis	D2	Report the statistical methods used for meta-analyses, if performed.
Additional analyses	16	Describe methods of additional analyses (eg, sensitivity or subgroup analyses, meta-regression), if done, indicating which were prespecified.
RESULTS		
Study selection	17	Provide numbers of studies screened, assessed for eligibility, included in the review (and included in meta-analysis, if applicable) with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics	18	For each included study provide citations and present key characteristics including: (a) participant characteristics (presentation, prior testing), (b) clinical setting, (c) study design, (d) target condition definition, (e) index test, (f) reference standard, (g) sample size, (h) funding sources
Risk of bias and applicability	19	Present evaluation of risk of bias and concerns regarding applicability for each study.
Results of individual studies	20	For each analysis in each study (eg unique combination of index test, reference standard, and positivity threshold) report 2 $\times$ 2 data (TP, FP, FN, TN) with estimates of diagnostic accuracy and confidence intervals, ideally with a forest or receiver operator characteristic (ROC) plot.
Synthesis of results	21	Describe test accuracy, including variability; if meta-analysis was done, include results and confidence intervals.
Additional analysis	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, adverse events).
DISCUSSION		
Summary of evidence	24	Summarize the main findings including the strength of evidence.
Limitations	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	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).
FUNDING		
Funding	27	For the systematic review, describe the sources of funding and other support and the role of the funders.

PRISMA flow diagram



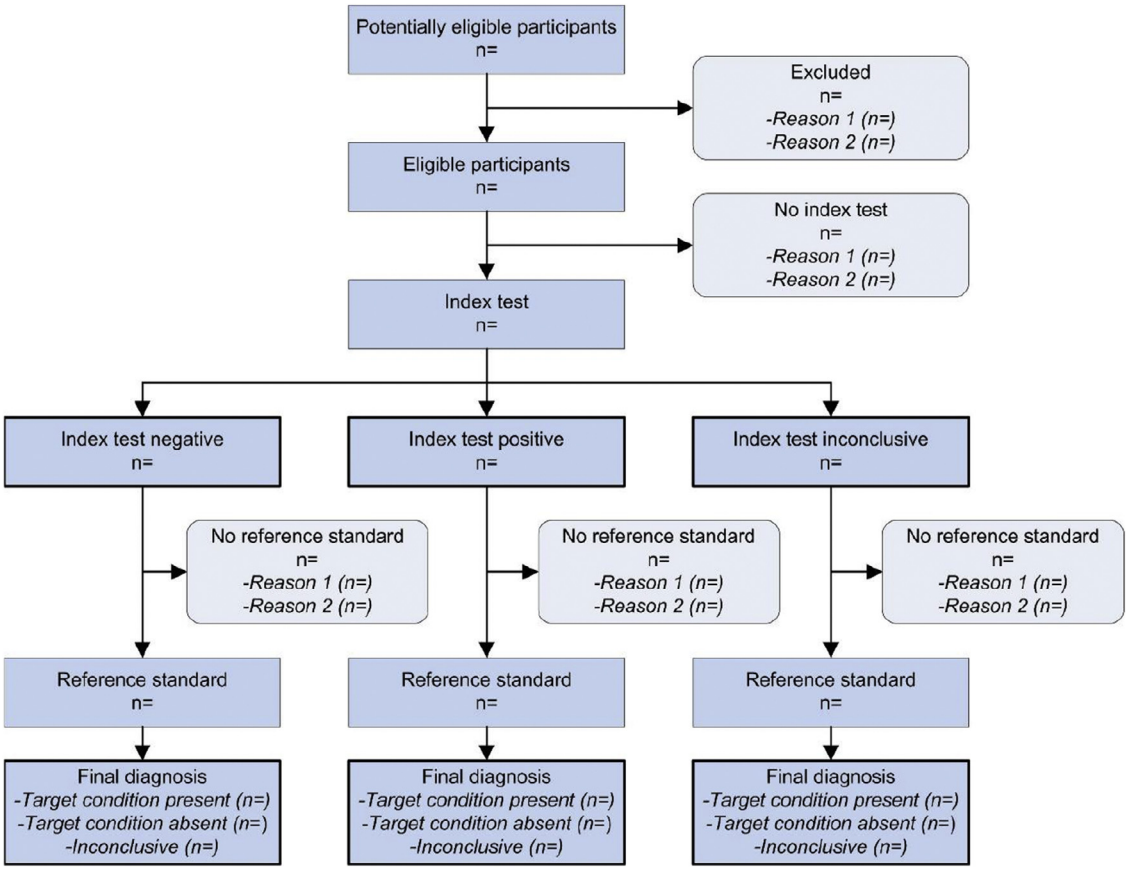
STARD 2015 checklist

Section/Topic	#	STARD 2015 Item
TITLE OR ABSTRACT		
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the index test
	4	Study objectives and hypotheses
METHODS		
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)
Participants	**6**	Eligibility criteria
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)
	8	Where and when potentially eligible participants were identified (setting, location and dates)
	9	Whether participants formed a consecutive, random or convenience series
Test methods	**10a**	Index test, in sufficient detail to allow replication
	10b	Reference standard, in sufficient detail to allow replication
	11	Rationale for choosing the reference standard (if alternatives exist)
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing prespecified from exploratory

(*Continued*)

Section/Topic	#	STARD 2015 Item
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing prespecified from exploratory
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test
	13b	Whether clinical information and index test results were available to the assessors of the reference standard
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy
	15	How indeterminate index test or reference standard results were handled
	16	How missing data on the index test and reference standard were handled
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory
	18	Intended sample size and how it was determined
RESULTS		
Participants	**19**	Flow of participants, using a diagram
	20	Baseline demographic and clinical characteristics of participants
	21a	Distribution of severity of disease in those with the target condition
	21b	Distribution of alternative diagnoses in those without the target condition
	22	Time interval and any clinical interventions between index test and reference standard
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	25	Any adverse events from performing the index test or the reference standard
DISCUSSION		
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalizability
	27	Implications for practice, including the intended use and clinical role of the index test
OTHER INFORMATION		
	28	Registration number and name of registry
	29	Where the full study protocol can be accessed
	30	Sources of funding and other support; role of funders

STARD 2015 flow diagram



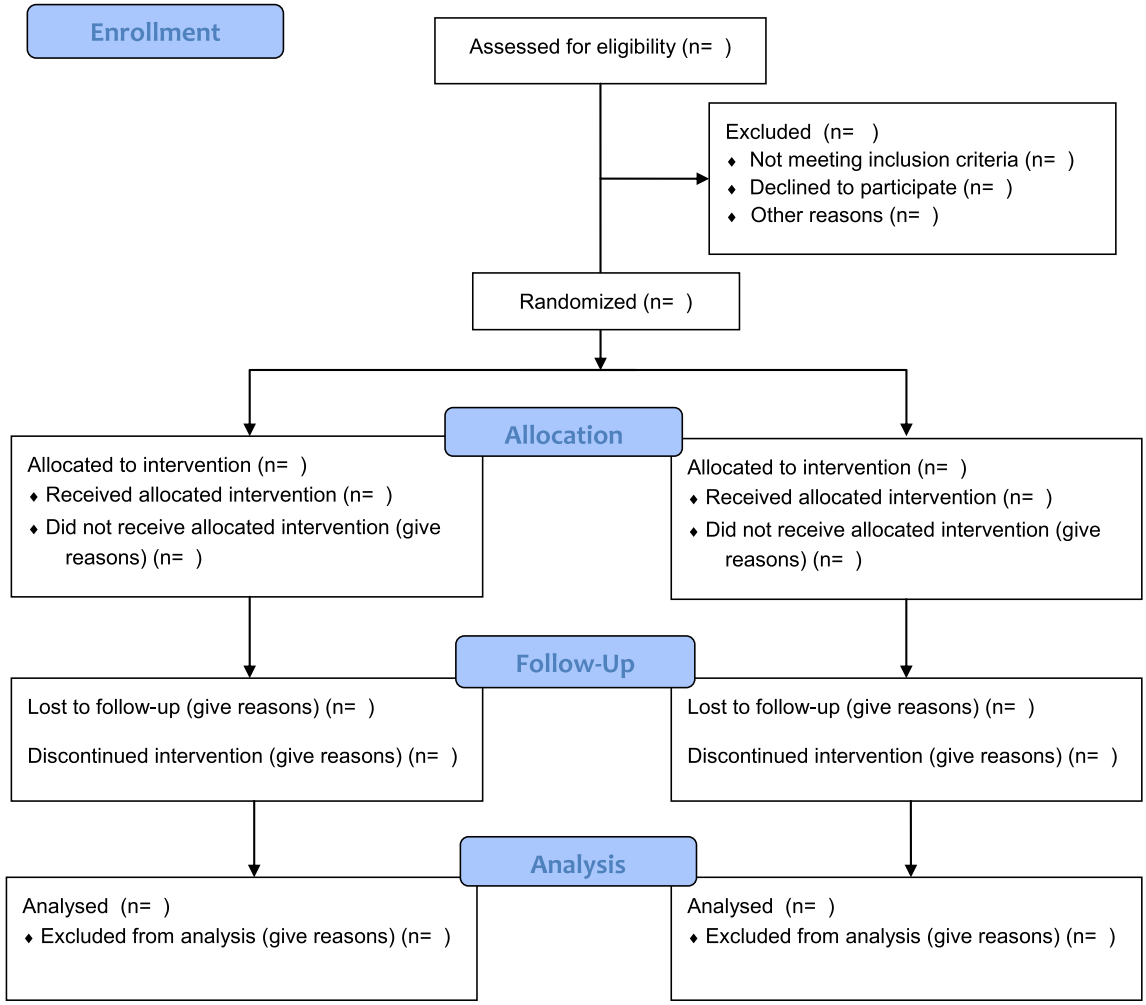
CONSORT checklist

Section/Topic	#	CONSORT 2010 Checklist item
Title and abstract		
	1a	Identification as a randomized trial in the title
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)
Introduction		
Background and objectives	2a	Scientific background and explanation of rationale
	2b	Specific objectives or hypotheses
Methods		
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons
Participants	4a	Eligibility criteria for participants
	4b	Settings and locations where the data were collected
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed
	6b	Any changes to trial outcomes after the trial commenced, with reasons
Sample size	7a	How sample size was determined
	7b	When applicable, explanation of any interim analyses and stopping guidelines
Randomization:		
Sequence generation	8a	Method used to generate the random allocation sequence
	8b	Type of randomization; details of any restriction (such as blocking and block size)
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how
	11b	If relevant, description of the similarity of interventions
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses
Results		
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome
	13b	For each group, losses and exclusions after randomization, together with reasons
Recruitment	14a	Dates defining the periods of recruitment and follow-up
	14b	Why the trial ended or was stopped
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group
Numbers analyzed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory
Harms	19	All-important harms or unintended effects in each group (for specific guidance see CONSORT for harms)

(*Continued*)

Section/Topic	#	CONSORT 2010 Checklist item
Discussion		
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses
Generalizability	21	Generalizability (external validity, applicability) of the trial findings
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence
Other information		
Registration	23	Registration number and name of trial registry
Protocol	24	Where the full trial protocol can be accessed, if available
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders

CONSORT 2010 flow diagram



References

1. Dwan K, Gamble C, Williamson PR, et al: Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. PLoS One 8:e66844, 2013
2. Sharifabadi AD, Korevaar DA, McGrath TA, et al: Reporting bias in imaging: Higher accuracy is linked to faster publication. Eur Radiol 28 (9):3632-3639, 2018. https://doi.org/10.1007/s00330-018-5354-x. [Epub 2018 Mar 21] PMID: 29564596
3. McGrath TA, McInnes MDF, van Es N, et al: Overinterpretation of research findings: Evidence of "spin" in systematic reviews of diagnostic accuracy studies. Clin Chem 63:1353-1362, 2017
4. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, et al: Helping doctors and patients make sense of health statistics. Psychol Sci Public Interest 8:53-96, 2007
5. Chowers MY, Gottesman BS, Leibovici L, et al: Reporting of adverse events in randomized controlled trials of highly active antiretroviral therapy: Systematic review. J Antimicrob Chemother 64:239-250, 2009
6. Chan AW, Altman DG: Epidemiology and reporting of randomised trials published in PubMed journals. Lancet 365:1159-1162, 2005
7. Tunis AS, McInnes MD, Hanna R, et al: 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 269:413-426, 2013
8. Ioannidis JP: The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q 94:485-514, 2016
9. Young NS, Ioannidis JP, Al-Ubaydli O: Why current publication practices may distort science. PLoS Med 5(10):e201, 2008
10. Moher D, Liberati A, Tetzlaff J, et al: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. BMJ 339: b2535, 2009
11. 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 319:388-396, 2018
12. Schulz KF, Altman DG, Moher D, et al: CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ 340:c332, 2010
13. Bossuyt PM, Reitsma JB, Bruns DE, et al: STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. BMJ 351: h5527, 2015
14. Plint AC, Moher D, Morrison A, et al: Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med J Aust 185:263-267, 2006
15. Smidt N, Rutjes AW, van der Windt DA, et al: The quality of diagnostic accuracy studies since the STARD statement: Has it improved? Neurology 67:792-797, 2006
16. 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 47:523-544, 2018
17. Moher D, Simera I, Schulz KF, et al: Helping editors, peer reviewers and authors improve the clarity, completeness and transparency of reporting health research. BMC Med 6:13, 2008
18. Equator Network. Reporting guidelines under development. Available from: http://www.equator-network.org/library/reporting-guidelines-under-development/
19. McGrath TA, McInnes MD, Korevaar DA, et al: Meta-analyses of diagnostic accuracy in imaging journals: Analysis of pooling techniques and their effect on summary estimates of diagnostic accuracy. Radiology 281:78-85, 2016
20. Zarei F, Zeinali-Rafsanjani B: Assessment of adherence of diagnostic accuracy studies published in radiology journals to STARD statement indexed in Web of Science, PubMed & Scopus in 2015. J Biomed Phys Eng 2018. *[E-Pub ahead of print]*
21. Roysri K, Chotipanich C, Laopaiboon V, et al: Quality assessment of research articles in nuclear medicine using STARD and QUADAS-2 Tools. Asia Ocean J Nucl Med Biol 2:120-126, 2014
22. Lodge MA, Lucas JD, Marsden PK, et al: A PET study of 18FDG uptake in soft tissue masses. Eur J Nucl Med 26:22-30, 1999
23. Whiting PF, Rutjes AW, Westwood ME, et al: A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. J Clin Epidemiol 66:1093-1104, 2013
24. Bossuyt PM, Reitsma JB, Bruns DE, et al: Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology 226:24-28, 2003
25. Cohen JF, Korevaar DA, Altman DG, et al: STARD 2015 guidelines for reporting diagnostic accuracy studies: Explanation and elaboration. BMJ Open 6:e012799, 2016
26. Korevaar DA, van Enst WA, Spijker R, et al: Reporting quality of diagnostic accuracy studies: A systematic review and meta-analysis of investigations on adherence to STARD. Evid Based Med 19:47-54, 2014
27. Bossuyt PM, Reitsma JB, Bruns DE, et al: STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. Clin Chem 61:1446-1452, 2015
28. Bossuyt PM, Reitsma JB, Bruns DE, et al: STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. Radiology 277:826-832, 2015
29. Bossuyt PMM: Chapter 19: STARD (STAndards for Reporting of Diagnostic Accuracy Studies). Guidelines for Reporting Health Research: A User's Manual. Wiley, 2014
30. Reuters T: Science citation index expanded-radiology, nuclear medicine & medical imaging-journal list. 2015. Available from: http://science.thomsonreuters.com.proxy.bib.uottawa.ca/cgi-bin/jrnlst/jlresults.cgi?PC=D&SC=VY
31. Frank R, McInnes M, Levine D, et al: Are Study and Journal Characteristics Reliable Indicators of "Truth" in Imaging Research? Radiology 287(1):215-223, 2018. https://doi.org/ 10.1148/radiol.2017170586. [Epub 2017 Nov 27] Review. PMID: 29173122
32. Bossuyt P. Google scholar profile. https://scholar.google.ca/citations?user=KTkjG9YAAAAJ&hl=en&oi=ao
33. Cuccurullo V, Cascini GL, Tamburrini O, et al: Bone metastases radiopharmaceuticals: an overview. Curr Radiopharm 6:41-47, 2013
34. Roberts CC, Daffner RH, Weissman BN, et al: ACR appropriateness criteria on metastatic bone disease. J Am Coll Radiol 7:400-409, 2010
35. O'Sullivan GJ, Carty FL, Cronin CG: Imaging of bone metastasis: An update. World J Radiol 7:202-211, 2015
36. Mehanna H, Wong WL, McConkey CC, et al: PET-CT surveillance versus neck dissection in advanced head and neck cancer. N Engl J Med 374:1444-1454, 2016
37. Rajarubendra N, Bolton D, Lawrentschuk N: Diagnosis of bone metastases in urological malignancies—an update. Urology 76:782-790, 2010
38. Brealey S, Scally AJ: Bias in plain film reading performance studies. Br J Radiol 74:307-316, 2001
39. Elmore JG, Wells CK, Lee CH, et al: Variability in radiologists' interpretations of mammograms. N Engl J Med 331:1493-1499, 1994
40. Whiting PF, Rutjes AW, Westwood ME, et al: QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529-536, 2011
41. Li ZZ, Huang YL, Song HJ, et al: The value of 18F-FDG-PET/CT in the diagnosis of solitary pulmonary nodules: A meta-analysis. Medicine (Baltimore) 97:e0130, 2018
42. Leeflang MM, Moons KG, Reitsma JB, et al: Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. Clin Chem 54:729-737, 2008
43. Justice AC, Covinsky KE, Berlin JA: Assessing the generalizability of prognostic information. Ann Intern Med 130:515-524, 1999
44. Knottnerus JA, Muris JW: Assessment of the accuracy of diagnostic tests: The cross-sectional study. J Clin Epidemiol 56:1118-1128, 2003
45. Kennedy-Dixon TG, Gossell-Williams M, Cooper M, et al: Evaluation of radiopharmaceutical adverse reaction reports to the british nuclear medicine society from 2007 to 2016. J Nucl Med 58:2010-2012, 2017
46. Pinto SR, Santos LFC, Dos Reis SRR, et al: Adverse reactions to radiopharmaceuticals: A survey based on clinical cases using criteria of systematic review. Ther Innov Regul Sci 52:109-113, 2018

47. Cohen J, Korevaar D, Gatsonis C, et al: STARD Group. STARD for abstracts: Essential items for reporting diagnostic accuracy studies in journal or conference abstracts. BMJ 358:j3751, 2017
48. Gardner IA, Nielsen SS, Whittington RJ, et al: Consensus-based reporting standards for diagnostic test accuracy studies for paratuberculosis in ruminants. Prev Vet Med 101:18-34, 2011
49. Noel-Storr AH, McCleery JM, Richard E, et al: Reporting standards for studies of diagnostic test accuracy in dementia: The STARDdem Initiative. Neurology 83:364-373, 2014
50. Kostoulas P, Nielsen SS, Branscum AJ, et al: STARD-BLCM: Standards for the reporting of diagnostic accuracy studies that use bayesian latent class models. Prev Vet Med 138:37-47, 2017
51. Higgins JPT, Green S: Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. Chichester, England; Hoboken, NJ: Wiley-Blackwell, 649, 2008
52. Moher D, Tetzlaff J, Tricco AC, et al: Epidemiology and reporting characteristics of systematic reviews. PLoS Med 4:e78, 2007
53. McInnes MD, Bossuyt PM: Pitfalls of systematic reviews and meta-analyses in imaging research. Radiology 277:13-21, 2015
54. Page MJ, Shamseer L, Altman DG, et al: Epidemiology and reporting characteristics of systematic reviews of biomedical research: A cross-sectional study. PLoS Med 13:e1002028, 2016
55. Alabousi M, Alabousi A, McGrath TA, et al: Epidemiology of systematic reviews in imaging journals: Evaluation of publication trends and sustainability? Eur Radiol 2018. https://doi.org/ 10.1007/s00330-018-5567-z. [Epub ahead of print] PMID: 30051140
56. McGrath T, McInnes M, Langer F, et al: Treatment of multiple test readers in diagnostic accuracy systematic reviews of imaging studies. Eur J Radiol 93:59-64, 2017
57. 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 6:194, 2017
58. Bossuyt PM, Irwig L, Craig J, et al: Comparative accuracy: Assessing new tests against existing diagnostic pathways. BMJ 332:1089-1092, 2006
59. Deeks J, Bossuyt P, Gatsonis C: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. The Cochrane Collaboration, 2013 1.0.0 ed
60. Reitsma JB, Glas AS, Rutjes AW, et al: Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J Clin Epidemiol 58:982-990, 2005
61. Rutter CM, Gatsonis CA: A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. Stat Med 20:2865-2884, 2001
62. Dong MJ, Zhao K, Lin XT, et al: Role of fluorodeoxyglucose-PET versus fluorodeoxyglucose-PET/computed tomography in detection of unknown primary tumor: A meta-analysis of the literature. Nucl Med Commun 29:791-802, 2008
63. Schulz KF, Altman DG, Moher D: CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ 340: c332, 2010
64. Moher D, Hopewell S, Schulz KF, et al: CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ 340:c869, 2010
65. Bothwell LE, Greene JA, Podolsky SH, et al: Assessing the Gold Standard—Lessons from the History of RCTs. N Engl J Med 374:2175-2181, 2016
66. Juni P, Altman DG, Egger M: Systematic reviews in health care: Assessing the quality of controlled clinical trials. BMJ 323:42-46, 2001
67. CONSORT. About CONSORT 2018 Available from: http://www.consort-statement.org/about-consort/history
68. Zanoni L, Bossert I, Matti A, et al: A review discussing fluciclovine ((18) F) PET/CT imaging in the detection of recurrent prostate cancer. Future Oncol 14:1101-1115, 2018
69. Akin-Akintayo OO, Jani AB, Odewole O, et al: Change in salvage radiotherapy management based on guidance with FACBC (Fluciclovine) PET/CT in postprostatectomy recurrent prostate cancer. Clin Nucl Med 42:e22-e28, 2017
70. Jani AB, Schreibmann E, Rossi PJ, et al: Impact of (18)F-fluciclovine PET on target volume definition for postprostatectomy salvage radiotherapy: Initial findings from a randomized trial. J Nucl Med 58:412-418, 2017
71. Lebech AM, Gaardsting A, Loft A, et al: Whole-Body (18)F-FDG PET/CT is superior to CT as first-line diagnostic imaging in patients referred with serious nonspecific symptoms or signs of cancer: A randomized prospective study of 200 patients. J Nucl Med 58:1058-1064, 2017
72. JNM. 2018. Available from: http://jnm.snmjournals.org
73. Radiology. Radiology—Publications. 2018. Available from:https://pubs.rsna.org/journal/radiology
74. Hopewell S, Dutton S, Yu LM, et al: The quality of reports of randomised trials in 2000 and 2006: Comparative study of articles indexed in PubMed. BMJ 340:c723, 2010
75. Egger M, Juni P, Bartlett C: Value of flow diagrams in reports of randomized controlled trials. JAMA 285:1996-1999, 2001
76. Gabriel SE, Normand SL: Getting the methods right—the foundation of patient-centered outcomes research. N Engl J Med 367:787-790, 2012
77. PCORI. Preliminary Draft Methodology Report - Patient-Centered Outcomes Research Institute. 2012. Available from: https://www.pcori.org/assets/Preliminary-Draft-Methodology-Report.pdf
78. Liu XT, Zhang X, Wen S, et al: Impact of the Consolidated Standards of Reporting Trials (CONSORT) checklist on reporting of randomized clinical trials in traditional Chinese medicine. J Evid Based Med 8:192-208, 2015
79. Ma B, Chen ZM, Xu JK, et al: Do the CONSORT and STRICTA checklists improve the reporting quality of acupuncture and moxibustion randomized controlled trials published in Chinese journals? A systematic review and analysis of trends. PLoS One 11:e0147244, 2016
80. Turner L, Shamseer L, Altman DG, et al: Does use of the CONSORT statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review. Syst Rev 1:60, 2012
81. Turner L, Shamseer L, Altman DG, et al: Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. Cochrane Database Syst Rev 11:Mr000030, 2012
82. Lassmann M, Chiesa C, Flux G, et al: EANM Dosimetry Committee guidance document: Good practice of clinical dosimetry reporting. Eur J Nucl Med Mol Imaging 38:192-200, 2011
83. Thoeny HC, Froehlich JM, Triantafyllou M, et al: Metastases in normal-sized pelvic lymph nodes: Detection with diffusion-weighted MR imaging. Radiology 273:125-135, 2014
84. deVet HCWEA, Riphagen II, Aertgeerts B, et al: Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy 0.4 ed. 2008. http://www.cochrane.org/editorial-and-publishing-policy-resource/cochrane-handbook-diagnostic-test-accuracy-reviews
85. Hutton B, Salanti G, Caldwell DM, et al: The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. Ann Intern Med 162:777-784, 2015
86. Stewart LA, Clarke M, Rovers M, et al: Preferred reporting items for systematic review and meta-analyses of individual participant data: The PRISMA-IPD statement. JAMA 313:1657-1665, 2015
87. Salameh JP, McInnes MDF, Moher D, et al: Completeness of Reporting of Systematic Reviews of Diagnostic Test Accuracy Based on the PRISMA-DTA Reporting Guideline. Clin Chem 2018. https://doi.org/ 10.1373/clinchem.2018.292987. [Epub ahead of print] PMID: 30237150
88. Samaan Z, Mbuagbaw L, Kosa D, et al: A systematic scoping review of adherence to reporting guidelines in health care literature. J Multidiscip Healthc 6:169-188, 2013
89. Hopewell S, Ravaud P, Baron G, et al: Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: Interrupted time series analysis. BMJ 344: e4178, 2012
90. van der Pol CB, McInnes MD, Petrcich W, et al: Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? PLoS ONE 10:e0119892, 2015
91. Dilauro M, McInnes MD, Korevaar DA, et al: Is there an association between STARD statement adherence and citation rate? Radiology 280:62-67, 2016