

STARD Adherence in an Interventional Radiology Guideline for Diagnostic Arteriography

Bryan Wright, BGS; Benjamin Howard, DO; Cole Wayant, BS; and Matt Vassar, PhD

Objective: To assess the quality of reporting in diagnostic accuracy studies (DAS) referenced by the Quality Improvement Guidelines for Diagnostic Arteriography and their adherence to the Standards for Reporting of Diagnostic Accuracy (STARD) statement.

Materials and Methods: Citations specific to the Society of Interventional Radiology's Quality Improvement Guidelines for Diagnostic Arteriography were collected. Using the 34-item STARD checklist, two authors in a duplicate and blinded fashion documented the number of items reported per diagnostic accuracy study. Authors met, and any discrepancies were resolved in a resolution meeting.

Results: Of the 26 diagnostic accuracy studies included, the mean number of STARD items reported was 17.8 (SD \pm 3.1). The median adherence was 18 (IQR, 17-19) items. Ten articles were published prior to 2003, the original date of STARD publication, and 16 articles were published after 2003. The mean number of reported items for the articles published prior to STARD 2003, and after STARD 2003 was 17.4 (SD \pm 2.4) and 18.1 (SD \pm 3.5), respectively. There were 14 STARD items that demonstrated an adherence of < 25%, and 13 an adherence > 75%.

Conclusion: The dichotomous distribution of adherence to the STARD statement by DAS investigated demonstrates that areas of deficient reporting may be present and require attention to ensure complete and transparent reporting in the future.

Keywords: Evidence-based medicine; Health care research; Research priorities

Health care models depend upon radiologists and interventional radiologists in their implementation of diagnostic imaging techniques to correctly identify numerous disease processes. However, the proper implementation of such diagnostic tests relies heavily on diagnostic accuracy studies (DAS), which are implemented to determine the ability of a diagnostic test to correctly distinguish between a disease and non-disease state of particular interest in individuals.¹ However, DAS are not devoid of factors that may prevent the proper implementation of such diagnostic techniques. For example, bias may be interjected into a DAS as a result of poor study design or failure to properly carry out a study.² Furthermore,

the ability of a test to correctly identify a pathologic condition of interest can be influenced by factors such as patient population, disease prevalence, and setting.³ Therefore, even when a study employs rigorous methods to limit bias, incomplete reporting of such factors may muddle clinical decision making for physicians. Furthermore, Ochodo and colleagues³ determined DAS specific to the field of radiology contain forms of spin. Other studies have also demonstrated that conclusions of DAS are frequently optimistic,⁴ and core methodological elements are commonly found to be omitted.⁵⁻⁷ Such findings may make DAS validation problematic without proper reporting within the field of medicine and radiology alike.

Corresponding Author: Bryan Wright, BGS, 1111 West 17th Street, Tulsa, OK 74107, Phone: 918-582-1972, Fax: 918-594-8449, Email: bdwright@okstate.edu

Statement of Data Access and Integrity: The authors declare that they had full access to all of the data in this study and the authors take complete responsibility for the integrity of the data and the accuracy of the data analysis.

Received: December 24, 2019
1st Revision: August 17, 2020
2nd Revision: November 13, 2020
Accepted: December 9, 2020

doi:10.3121/cmr.2020.1549

To address incomplete reporting in DAS, improve transparency, and provide the public a method to evaluate published manuscripts, the Standards for Reporting of Diagnostic Accuracy (STARD) statement was developed in 2003.⁸ STARD was initially developed to provide an applicable checklist for DAS similar to that of the Consolidated Standards for the Reporting of Trials. Since then, the STARD statement has been updated to incorporate new evidence about bias and interpretation of DAS, as well as to make the STARD checklist easier to use.⁹ Accompanying the newest STARD statement is a checklist consisting of 34 items essential to the complete reporting of a DAS. These items range from the structure of titles and abstracts through study limitations. Each item is essential in its own right, either for the sake of improving clinical decision making, ensuring reproducibility of the study, transparency, or mitigating research waste.

In this investigation we applied the STARD statement and checklist to DAS used as evidence in the construction of the Society of Interventional Radiology's (SIR) Quality Improvement Guideline for Diagnostic Arteriography.¹⁰ The guideline covers a wide range of diagnostic tests across many organ systems, making it an important resource for physicians. The primary objective of the investigation was to assess the quality of reporting in DAS referenced in the SIR Quality Improvement Guidelines for Diagnostic Arteriography using the STARD statement and checklist. The exploratory objectives were to identify STARD checklist items with < 25% adherence and to compare DAS published before and after 2003, corresponding to the original date of STARD publication.

Methods

For our investigation, the Quality Improvement Guidelines for Diagnostic Arteriography from the SIR were selected.¹¹ Quality improvement guidelines, as defined by the SIR, are evidence-based documents developed through a systematic consensus that seek to improve the standard of care by reducing the variations witnessed in clinical decision making.¹² The Diagnostic Arteriography guideline was selected due to its extensive scope and clinical relevance, as it covered topics ranging from catheter-based arteriography to the noninvasive cross-sectional modalities such as multi-detector computed tomography angiography and magnetic resonance angiography that have largely replaced catheter arteriography. All references from the guidelines were added to a PubMed collection and subsequently exported into a collection in Rayyan,¹³ which is a web-based platform designed to increase the efficiency and accuracy of screening references. All references were screened for inclusion by TT and BH in a duplicate and blinded fashion. All discrepancies were resolved by group discussion.

The Cochrane Handbook for Diagnostic Test Accuracy Reviews¹⁴ and the STARD statement⁹ for the description and

identification of a DAS were adhered to strictly. Cochrane defines DAS as an investigation to “obtain how well a test, or a series of tests, is able to correctly identify diseased patients or, more generally, patients with the target condition.”¹⁴ According to STARD, at least one measure of accuracy (sensitivity, specificity, predictive values, etc.) is necessary to identify a study of diagnostic accuracy.⁴ To qualify for inclusion as a DAS, all articles had to meet this definition. Key exclusion criteria were: clinical trials, observational studies, clinical guidelines, narrative reviews, meta-analyses, and any other citations not meeting the definition of a DAS.

Each included study was assessed according to the STARD statement. The statement is accompanied by a 34-item checklist that can be used to ensure the complete reporting of all aspects of DAS. The STARD checklist items are categorized in seven major sections: Title/Abstract, Abstract, Introduction, Methods, Results, Discussion, and Other Information. We devised a Google Form based on the STARD 2015 checklist and pilot tested it for calibration. During calibration, both data extractors (TT and BH) referred to the STARD 2015 Explanation and Elaboration¹⁵ and the STARD for abstracts¹⁶ documents to ensure a complete understanding of all STARD checklist items. The STARD items were assigned a “yes” for adherence or a “no” for non-adherence. However, items 13a and 13b were reworded after our pilot test. Items 13a and 13b pertained to whether a patient's clinical information was available to the study investigators, and answering “yes” is indicative of poor research practices. We reworded this question to ensure that a “yes” equated to good research practices, which is consistent with the other STARD items. TT and BH were blinded to each other's responses during data extraction. Discrepancies were resolved by group discussion. Once a consensus was reached, the data were exported to a Google Sheet to calculate summary statistics.

Results

The SIR guideline contained 128 citations, from which 14 DAS were identified. From the remaining citations, all narrative reviews were separated and analyzed—defined as a review without a systematic database search—in an attempt to identify additional DAS for inclusion. From the narrative reviews, 12 more unique DAS were identified, giving a total sample of 26 DAS. A flow diagram with included and excluded studies is shown in Figure 1.

Out of 34 STARD items, a mean of 17.8 (SD ± 3.1) items were adhered to. The median adherence was 18 (IQR, 17-19) items. We found that 14 of the 34 STARD checklist items had an adherence of < 25%: one item in the “Abstract” section, six in the “Methods” section, four in the “Results” section, and three in the “Other Information” section. Conversely, we found that 13 of the 34 items demonstrated an adherence of ≥ 75%: one item in the “Title/Abstract” section, two in the “Introduction” section, eight in the “Methods” section, two in

the “Results” section, and one in the “Discussion” section. The overall adherence to each STARD item is shown in Table 1.

Key items with < 25% adherence were as follows: authors failed to provide a structured and informative abstract (Item 2) that was compliant with the STARD statement, specifically the STARD for Abstracts checklist;¹⁶ a rationale for the choice of reference standard (Item 11); a distribution of alternative diagnoses in those without the target condition (Item 21b); a time interval between index and reference standard and if any clinical interventions occurred in that time interval (Item 22); whether any adverse events resulted as a consequence of performing the index test or reference standard (Item 25); the registration number and/or name of the study registry (Item 28); the location where the full study protocol could be accessed (Item 29); and the source of funding for the study (Item 30).

Conversely, several key items had adherence > 75% and were as follows: the basis (such as symptoms or previous tests results) on which potentially eligible patients were identified (Item 7); a sufficiently detailed explanation of the index test to allow for replication (Item 10a); blinding of test performers and/or readers to clinical information and the results of previous tests (Items 13a,13b); and the baseline demographics and clinical features of study participants (Item 20).

Of the articles, 10 were published before 2003 (range 1990-2001), corresponding to the original date of STARD publication, and 16 articles were published after 2003 (range 2006-2013). The mean number of reported items for the articles published prior to STARD 2003 was 17.4 (SD ± 2.4) and 18.1 (SD ± 3.5) for items published after STARD 2003.

Discussion

Overall, the results demonstrate that DAS used in support of the SIR diagnostic arteriography guideline may have room for improvement within reporting practices. Namely, there was a distinct dichotomy between the majority of items—adherence was either extremely high or low. Of the items, 14 showed adherence of < 25%, and 13 showed adherence > 75%. Such a finding has implications for clinical practice and guideline recommendations, which may represent research inertia with respect to improving reporting practices.

Our results are consistent with a recent comparison of STARD adherence between DAS published in 2000, 2004, and 2012.⁶ The study found improvements over time in adherence to STARD items, but concluded that the improvements were smaller than ideal. For example, both studies found that DAS often still fails to report eligibility criteria for patients and time interval between index test and reference standard. However, our investigation highlights some key, persistently under-reported STARD items. First, authors infrequently provided a rationale for their choice of reference standard. Sometimes, multiple reference standards are available, while other times only a few are available. In any case, authors frequently neglected to provide a rationale for their choice of test or stated that no alternative reference standard was available. The comparison of the index test depends on a proper selection of a reference standard. Practical or ethical factors may play a role in choosing a reference standard, and these factors are important information for clinical practice. Second, authors often failed to indicate how much time passed between the index test and reference standard and whether any clinical interventions took place during this time. Since DAS are cross-sectional investigations, a delay between the two tests may bias the interpretation, result, and comparison of the tests. Therefore, to interpret and apply the results of a DAS, it

is imperative for the reader to know the point(s) in time the two tests were performed. Third, authors frequently failed to mention their funding source, leaving the reader blind to an important form of bias that is common across the medical literature.¹⁷

As often as key items were under-reported, there were items that were well reported. For example, authors commonly discussed blinding procedures, provided patient demographics, and discussed implications for practice. These three items, and more, are essential components for the application of DAS in clinical practice, but their applicability is

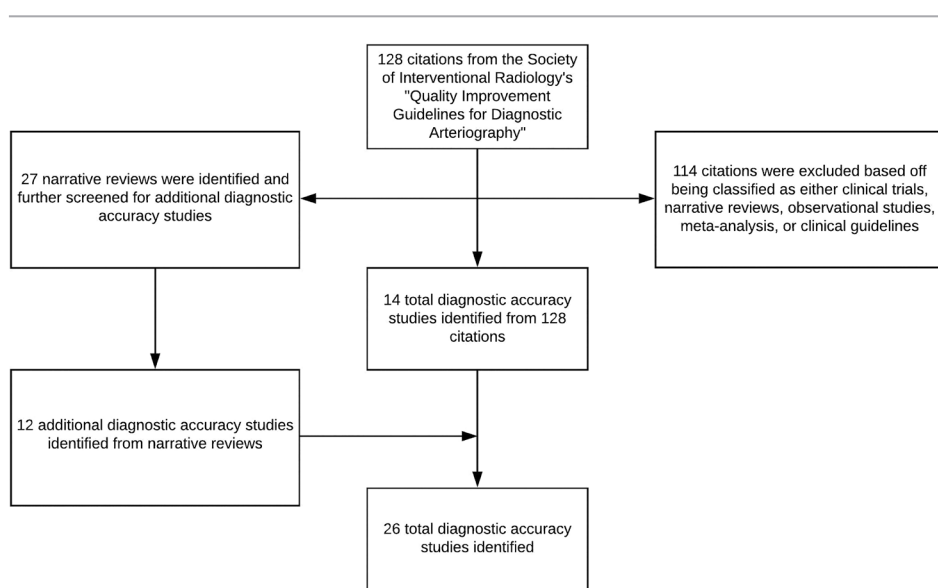


Figure 1. Study inclusion and exclusion process

Table 1: Item Specific Adherence to STARD

Category and Item No.	No. of articles reporting the item (% adherence) n=26
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)	23 (88.5%)
Abstract	
2. Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	0 (0%)
Introduction	
3. Scientific and clinical background, including the intended use and clinical role of the index test	26 (100%)
4. Study objectives and hypotheses	25 (96.2%)
Methods	
<i>Study design</i>	
5. Whether data collection was prospective study or retrospective study	26 (100%)
<i>Participants</i>	
6. Eligibility criteria	15 (57.7%)
7. How potentially eligible participants were identified (such as symptoms, results from previous tests)	26 (100%)
8. Where and when potentially eligible participants were identified (setting, location and dates)	4 (15.4%)
9. Whether participants formed a consecutive, random or convenience series	26 (100%)
<i>Test methods</i>	
10a. Index test, in sufficient detail to allow replication	26 (100%)
10b. Reference standard, in sufficient detail to allow replication	19 (73.1%)
11. Rationale for choosing the reference standard (if alternatives exist)	2 (7.7%)
12a. Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	16 (61.5%)
12b. Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	12 (46.2%)
13a. Whether the performers/readers of the index test were blinded to clinical information and reference standard results	21 (80.8%)
13b. Whether the assessors of the reference standard were blinded to clinical information and index test	23 (88.5%)
<i>Analysis</i>	
14. Methods for estimating or comparing measures of diagnostic accuracy	20 (76.9%)
15. How indeterminate index test or reference standard results were handled	5 (19.2%)
16. How missing data on the index test and reference standard were handled	2 (7.7%)
17. Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	3 (11.5%)
18. Intended sample size and how it was determined	0 (0%)
Results	
<i>Participants</i>	
19. Flow of participants, using a diagram	5 (19.2%)
20. Baseline demographic and clinical characteristics of participants	25 (96.2%)
21a. Distribution of severity of disease in those with the target condition	18 (69.2%)
21b. Distribution of alternative diagnoses in those without the target condition	1 (3.9%)
22. Time interval and any clinical interventions between index test and reference standard	1 (3.9%)
<i>Test results</i>	
23. Cross tabulation of the index test results (or their distribution) by the results of the reference standard	18 (69.2%)
24. Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	23 (88.5%)
25. Any adverse events from performing the index test or the reference standard	6 (23.1%)
Discussion	
26. Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	18 (69.2%)
27. Implications for practice, including the intended use and clinical role of the index test	26 (100%)
Other Information	
28. Registration number and name of registry	0 (0%)
29. Where the full study protocol can be accessed	0 (0%)
30. Sources of funding and other support; role of funders	2 (7.7%)

hindered if the aforementioned under-reported items are absent. For example, despite assessors being blinded, if the assessments of the index test and reference standard were separated in time, the results of either may be skewed. Additionally, if an inadequate reference standard was chosen, and no discussion of other available tests was provided, the performance of the index test in comparison may be biased. Lastly, if financial conflicts of interest are not reported, there remains an underlying suspicion for bias in the interpretation and presentation of study findings.

Strengths and Limitations

Regarding strengths, we sampled a selection of DAS specific to a guideline applicable to diagnostic testing across various organ systems. Such an investigation ensures results are not broadly generalized to numerous guidelines, but instead are specific to the SIR diagnostic arteriography guideline. Furthermore, our methodology for data extraction was conducted in a similar fashion to systematic reviews by following the *Cochrane Handbook*.¹⁸ Regarding study limitations, our results may not be generalized to all DAS studies, since our sample was taken from one guideline. Another potential limitation of our study was that we weighted all items equally in the production of the summary score. It could be argued that certain items are more clinically meaningful than others, or that certain items have a greater evidence base to support their inclusion. Future studies should evaluate the utility of a summary score for STARD, such as the Yamato et al¹⁹ study on the TIDieR Checklist. Furthermore, our study was limited to analyzing mostly single-site studies as only one multi-centered study was included. Thus, a comparison between the two could not be made.

Conclusion

In summation, we found dichotomy in the adherence to STARD items in the DAS used in support of the SIR Quality Improvement Guideline for Diagnostic Arteriography. Out of 34 STARD items, 14 had < 25% adherence, and 13 had > 75% adherence. For the sake of progress in evidence-based clinical decision making, the quality of reporting in future DAS must improve and guideline developers must take into account the deficiencies in past DAS reporting. We commend the SIR for their stringent methods of identifying and appraising the diagnostic accuracy literature, but we recommend the application of the STARD checklist when assessing the quality of the included evidence.

References

- Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. *BMJ*. 2012;345(jul 02 1):e3999.
- Kim KW, Lee J, Choi SH, Huh J, Park SH. Systematic Review and Meta-Analysis of Studies Evaluating Diagnostic Test Accuracy: A Practical Review for Clinical Researchers-Part I. General Guidance and Tips. *Korean J Radiol*. 2015;16(6):1175-1187.
- Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeflang MM. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of “spin”. *Radiology*. 2013;267(2):581-588.
- Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
- Korevaar DA, van Enst WA, Spijker R, Bossuyt PM, Hooft L. Reporting quality of diagnostic accuracy studies: a systematic review and meta-analysis of investigations on adherence to STARD. *Evid Based Med*. 2014;19(2):47-54.
- Korevaar DA, Wang J, van Enst WA, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. *Radiology*. 2015;274(3):781-789.
- Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282(11):1061-1066.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Radiology*. 2003;226(1):24-28.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351:h5527.
- Dariushnia SR, Gill AE, Martin LG, et al. Quality improvement guidelines for diagnostic arteriography. *J Vasc Interv Radiol*. 2014;25(12):1873-1881.
- Dariushnia SR, Gill AE, Martin LG, et al. Quality improvement guidelines for diagnostic arteriography. *J Vasc Interv Radiol*. 2014;25(12):1873-1881.
- Society of Interventional Radiology- Standards development process [Internet]. Available from: <https://www.sirweb.org/practice-resources/clinical-practice/guidelines-and-statements/methodology/>. Last accessed March 10, 2021.
- Ouzzani M, Hammady H, Fedorowicz, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*. 2016;5:210.
- Bossuyt PM, Leeflang MM. Chapter 6: Developing criteria for including studies. In: *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 0 4* [updated September 2008] The Cochrane Collaboration. 2008
- Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6(11):e012799.
- Cohen JF, Korevaar DA, Gatsonis CA, et al. STARD for Abstracts: essential items for reporting diagnostic accuracy studies in journal or conference abstracts. *BMJ*. 2017;358:j3751.

17. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database Syst Rev.* 2017;2:MR000033.
18. *Cochrane Handbook for Systematic Reviews of Interventions* [Internet]. Higgins J, Green S, eds. Version 5.1.0 [updated March 2011]. Available from: https://handbook-5-1.cochrane.org/index.htm#chapter_7/7_4_3_individual_patient_data.htm. Accessed Jun 25, 2019.
19. Yamato T, Maher C, Saragiotto B, et al. The TIDieR Checklist Will Benefit the Physiotherapy Profession. *Physiother Can.* 2016;68(4):311-314.

Author Affiliations

Bryan Wright, BGS; Benjamin Howard, DO*; Cole Wayant, BS*; and Matt Vassar, PhD**

**Oklahoma State University Center for Health Sciences, Psychology Department, 1111 West 17th Street, Tulsa, OK 74107*