World Journal of Urology (2019) 37:969–974
https://doi.org/10.1007/s00345-018-2446-9

ORIGINAL ARTICLE

The reporting quality of studies of diagnostic accuracy in the urologic literature

Daniel W. Smith[1,2] · **Shreyas Gandhi**[3] · **Philipp Dahm**[1,2]

Received: 12 May 2018 / Accepted: 11 August 2018 / Published online: 23 August 2018

Abstract

Introduction High-quality evidence regarding questions of diagnostic accuracy relies on transparent reporting of study results. The quality of reporting for such studies in the urologic literature is unknown.

Methods In accordance with an a priori protocol, we systematically searched for all articles on diagnostic accuracy studies published in four major urologic journals in 2015. Using the 2015 STAndards for Reporting Diagnostic accuracy studies (STARD) checklist, two of us independently abstracted data. For each article, we calculated STARD summary scores (scale of 0–30, with higher scores reflecting higher-quality reporting). We compared scores by journal, topic, and sample size.

Results We screened 819 references of which 61 met inclusion criteria. Nearly two-thirds of studies (39/61%; 63.9%) addressed prostate cancer diagnosis or staging; less than one in ten (6/61%; 9.8%) was conducted in non-oncological disease settings. The major focus for the investigation of new index tests lay in imaging modalities (33/61%; 54.1%); over half of these imaging studies addressed magnetic resonance imaging (18/61%; 29.5%). The average STARD score was 18.9 ± 2.4 (range 12–24). Six criteria had poor reporting compliance and were met by less than 20% of studies. We found no association between reporting quality and topic, journal or study size.

Conclusions The reporting quality of studies of diagnostic accuracy appears modest and independent of topic, journal or study size. There is an urgent need for greater awareness for the reporting quality of these studies among readers, editors, and investigators to raise evidentiary standards on issues of diagnosis.

Keywords Diagnostic test accuracy · Reporting quality · STARD criteria

Introduction

Evidence-based health care relies on methodologically rigorous and transparently reported research evidence. The Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network is a not-for-profit organization that provides a depository for such reporting guidelines; for randomized controlled trials, the CONsolidated Standards Of Reporting Trials (CONSORT) are among the criteria most well-known to urologists [1, 2]. A recent study suggested that randomized controlled trial reporting has improved over recent years; this would appear to indicate greater awareness of CONSORT criteria among study authors, journal editors, and readers. As a result, reported evidence is improved and better suited to support clinical and health policy decision-making [3].

The STAndards for Reporting Diagnostic accuracy studies (STARD) checklist is a corresponding instrument for articles on diagnostic accuracy studies. Its objective is to promote transparent and complete reporting of such studies to allow readers to critically appraise them for their validity, impact and applicability, and to assess for bias and generalizability [4, 5]. Though initially published in 2003 and updated in 2015, the STARD checklist is much less widely known than the CONSORT criteria or many other EQUATOR

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00345-018-2446-9) contains supplementary material, which is available to authorized users.

✉ Philipp Dahm
pdahm@umn.edu

[1] Minneapolis VA Health Care System, Urology Section 112D, University of Minnesota, One Veterans Drive, Minneapolis, MN 55417, USA

[2] Department of Urology, University of Minnesota, Minneapolis, MN, USA

[3] Department of Urology, Dalhousie University, Halifax, Nova Scotia, Canada

Network guidelines [4, 6]. The updated STARD checklist is divided into seven categories: title, abstract, introduction, methods, results, discussion, and other information (see Online Appendix for full STARD checklist). To date, no study (until ours) has used the STARD checklist to formally assess reporting quality of articles on diagnostic accuracy studies in urology: an important deficit, given the increasing role of diagnostic tests in the practice of urology. In this study, we formally assessed the reporting quality of articles on diagnostic accuracy studies published in four major urologic journals in 2015 to better understand compliance and to what extent they provide the necessary details to convince critical readers that their results are trustworthy, clinically meaningful and relevant to patient care.

Methods

In accordance with an a priori protocol (Research Registry #2214), we performed a PubMed search of four major urologic journals: *Journal of Urology*, *European Urology*, *BJU International*, and *Urology*. The aforementioned are the urology journals with the highest impact factor and have in the past been the focus of similar studies assessing methodological and/or reporting quality or the urological literature [3, 7, 8]. To find articles published in 2015 on diagnostic accuracy studies, we used the Clinical Queries filter for diagnosis and set it on "broad" [9]. An additional table of contents search was undertaken to ensure completeness. We screened our search results via a two-tiered process of (1) the title/abstracts and (2) the full text, using Covidence systematic review software [10].

From each article, two of us (DWS, SG) independently abstracted data. We applied the 2015 STARD [6] checklist, an update of the original 2003 version [4]. To assure clear, consistent interpretation of STARD checklist items, we had previously pilot-tested that form in two sets of three studies. For each article, we scored each STARD criterion as met (one point), not met (zero points), or not applicable (N/A). In addition, when a specific STARD criterion entailed two distinct subcategories, we gave articles half a point if they met one of those two subcategories; one point, if they met both subcategories.

Then, the two of us (DWS, SG) entered each criterion and corresponding scores onto a separate electronic spreadsheet. All three of us reviewed any discrepancies and resolved them by discussion and consensus. We not only assessed adherence to STARD criteria but also further characterized the studies by sample size, funding source, type of data collection (prospective vs. retrospective), patient population (pediatric vs. adult), topic within urology, and type of diagnostic test (imaging modality, biomarker, tissue sampling and/or biopsy, and other).

The primary objective of this study was to assess adherence to STARD criteria within the urologic literature for the year 2015. For our descriptive statistical analysis, we used SPSS version 26.0. For each article, going by the number of STARD criteria met per article, we calculated a STARD summary score on a scale of 0–30 (with higher scores reflecting higher-quality reporting). Individual STARD criteria are reported herein as proportions and percentages. To assess for differences in reporting by journal of publication, topic, and by sample size, we performed statistical hypothesis-testing using the student's *t* test and ANOVA. The protocol for our methodologic study was not eligible for registration with PROSPERO, the international prospective register of systematic reviews, but we instead prospectively registered it with the Research Registry (#2214) in February 2017 [11].

Results

Our literature search identified 819 articles, which we then screened on the basis of their titles and abstracts. After excluding 752 articles, we obtained the full text of 67, of which 61 met our study inclusion criteria (Fig. 1). The reasons we excluded six articles at the full-text stage were as follows: three articles described studies that were systematic reviews and/or meta-analyses; one article was an editorial response; one article described an investigational study of ultrasound and bladder outlet obstruction (BOO) models; and one article described an animal study of bladder permeability.

Of the 61 articles, 60 (98.4%)—the overwhelming majority—described diagnostic accuracy studies in adult patients; only one, in pediatric patients (Table 1). Of the 61 articles, 39 (63.9%)—nearly two-thirds— addressed prostate cancer diagnosis or staging. Of the remaining studies addressing oncology, eight studies (13%) focused on urothelial cell cancer, five (8%) on renal cell cancer, and three (5%) on penile cancer. Only six (9.8%) involved patients with non-oncologic diseases, namely voiding dysfunction, three (4.7%); stone disease, two (3.2%); and sexual dysfunction, one (1.6%). The journals of publication were the *Journal of Urology* ($n=23$), *BJU International* ($n=21$), *Urology* ($n=10$) and *European Urology* ($n=7$).

With regard to new index tests, 33 (54.1%) of the articles focused on imaging modalities. Of those 33 articles, 18 (29.5%)—slightly more than half—addressed magnetic resonance imaging (MRI). Only 14 (23%) articles—just under one-quarter—assessed biomarkers as index tests; only nine (15%), tissue sampling and/or biopsy. With regard to type of data collection, 38 (61%) of the 61 articles described prospective studies. The median sample size was 154 patients (interquartile range,

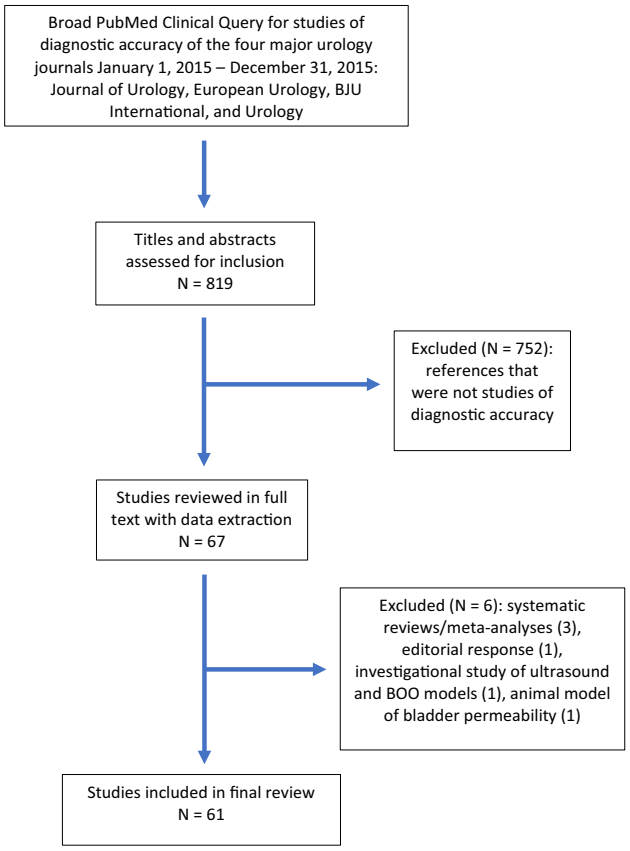

Fig. 1 Flowchart of literature review, abstract assessment, and full-text data abstraction

93–353). Only 27 (44%) of the 61 articles—fewer than half of them— reported a study funding source. Overall, 12 (19%) of the 61 articles acknowledged industry funding; the remainder either indicated no funding source or reported studies that had been funded by their institution or the government.

The total number of individual STARD criteria (including subcategories) was 34. The median (interquartile range) of kappa values as a measure of interobserver agreement beyond chance across all 34 items was 0.59 (0.28; 1.0) and the mean (± standard deviation) was 0.59 ± 0.36, which reflects moderate agreement [12]. Reporting compliance among the 61 articles ranged widely by individual criteria (from 4.9 to 100%). It reached 100% for several individual criteria, such as #3 (scientific & clinical background), #10a (index test: sufficient detail for replication), and #27 (implications for practice). However, reporting compliance was a mere 4.9% for both #18 (sample size determination) and #28 (registration number and name of registry).

Table 1 Study characteristics ($n = 61$)

Study characteristics	*N* (%)
Articles by journal	
BJU International	21 (34.4%)
Eur Urol	7 (11.5)
JU	23 (37.7%)
Urology	10 (16.4%)
Topic within urology, *n* (%)	
Prostate cancer	39 (63.9%)
Renal cell cancer	5 (8%)
Urothelial cell cancer	8 (13%)
Penile cancer	3 (5%)
Non-oncologic diseases	6 (9.8%)
Patient population, *n* (%)	
Adult	60 (98.4%)
Pediatric	1 (1.6%)
Type of data collection, *n* (%)	
Prospective	38 (61%)
Retrospective	24 (39%)
Sample size, *n* (%)	
150 or fewer patients	29 (48%)
More than 150 patients	32 (52%)
Type of diagnostic test, *n* (%)	
Imaging modality	33 (54.1%)
Biomarker	14 (23%)
Tissue sampling and/or biopsy	9 (15%)
Other	5 (8%)
Funding source, *n* (%)	
Industry	12 (19%)
Institution	4 (7%)
Government	8 (13%)
No funding source	3 (5%)
No information provided	34 (56%)

For 16 of the 34 STARD criteria, reporting compliance reached at least 80% (Fig. 2). But for six criteria, reporting compliance was less than 20%. Poor reporting quality was most notable in the discussion section of articles for criteria #15 (handling of intermediate results), #16 (handling of missing data), #17 (pre-specified or exploratory analysis), #18 (sample size determination), #28 (registration number and name of registration), and #29 (location of full study protocol).

The mean number of STARD criteria met was 18.9 ± 2.4 (range 12–24); the median number was 19.5 (interquartile range 17.3–20.5). In a univariate analysis, we found no statistically significant association between reporting quality and any of these three variables: topic within urology, sample size, or journal of publication (Table 2).

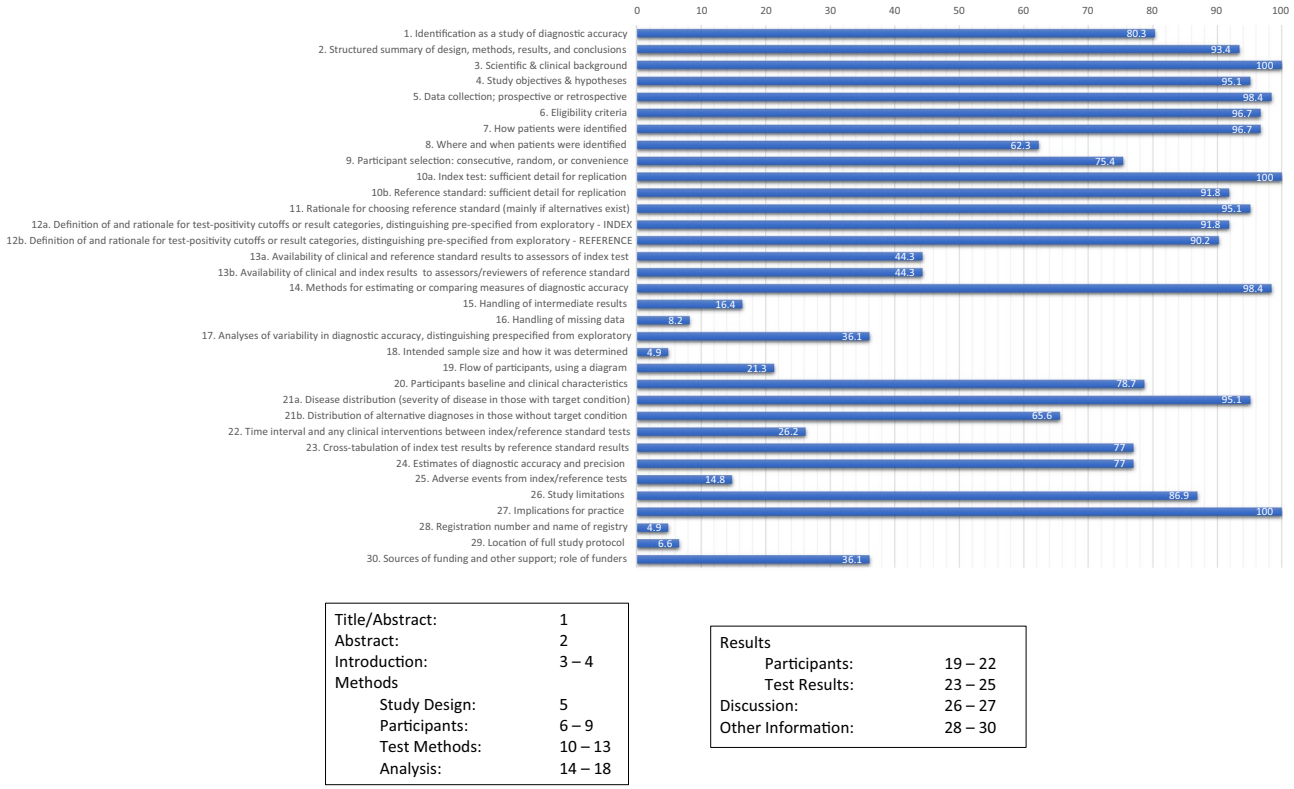

Fig. 2 Individual STARD criteria met by percentage

Table 2 Association between STARD scores and Journal of Publication, Topic, and Sample Size

	STARD score (SD)	*P*-value
Journal of publication		
BJU International	20.0 (2.3)	0.052
European Urology	19.0 (2.9)	
Journal of Urology	18.7 (1.7)	
Urology	17.6 (2.8)	
Topic		
Prostate cancer	18.9 (2.5)	0.659
Renal cell cancer	18.4 (1.9)	
Urothelial cell cancer	18.9 (2.8)	
Penile cancer	21.0 (2.8)	
Non-oncologic diseases	19.1 (1.2)	
Sample size		
150 or fewer patients	18.6 (2.1)	0.20
More than 150 patients	19.4 (2.6)	

STARD STAndards for Reporting Diagnostic accuracy studies, *SD* standard deviation

Discussion

In this study, our principal finding was the major heterogeneity, with regard to compliance with individual STARD criteria, of articles published in 2015 on diagnostic accuracy studies in urology. Many criteria were reported well, especially those related to the title, abstract, introduction, and methods. But other criteria were poorly addressed.

The worst reporting deficits (as evidenced by a reporting compliance rate of less than 20% of studies) involved handling of indeterminate results and missing data, sample size determination, registration number and name of registration, and location of full study protocol. Each of those reporting deficits has a corresponding CONSORT criterion. Currently, the CONSORT checklist has been broadly endorsed in the general medical and urologic literature [3]. For example, articles on randomized controlled trials now routinely provide information on pre-trial sample size estimates. Most reputable journals will not publish results of trials that have not been registered in a clinical trial registry. Such registration requires a trial protocol that outlines all pertinent aspects of the study, including the basic design, the study endpoints, the frequency and method of follow-up, and the analytic approach. Those issues are equally important in articles on diagnostic accuracy studies. That information is, in fact, critical, enabling clinicians, guideline developers, and health care policymakers to assess the validity of studies and thus the reliability of their results [5]. Our findings appear particularly relevant to the field of prostate cancer imaging (especially MRI)—a rapidly evolving area of research that

contributed one-third of the articles we analyzed. But many of those articles did not provide the necessary methodologic detail for readers to assess the validity of the described studies. In addition, it appears noteworthy that less than one in five publications acknowledged industry funding, whereas it might be expected that a majority of studies related to new imaging modalities or laboratory tests involve industry support. Greater emphasis on prospective study registration would enhance transparency in terms of relevant conflict of interest disclosures.

The strengths of this study include our use of the STARD checklist, which provides well-established, standardized criteria to assess reporting quality of articles on diagnostic accuracy studies. Our research was governed by a registered, a priori written protocol. We conducted a systematic literature search and two members of our investigative team independently screened and abstracted data. Our study is the first of its kind to assess the urologic literature. The limitations of this study include our focus on studies published in the four main urologic journals. We recognize that many diagnostic accuracy studies relevant to the practice of urology are instead published in other, often high-impact non-urologic journals, such as the PROState Magnetic Resonance Imaging Study (PROMIS) on the use of prostate MRI to guide prostate biopsies [13]. However, a comparison of reporting quality of urologic vs. non-urologic journals was outside the scope of this research project. Second, we acknowledge the limitation of any scoring system of reporting quality and methodologic quality; any such system assumes that individual criteria deserve equal weight. Nonetheless, we believe a summary score provides a useful idea of the gestalt of overall reporting quality. Last, our study sample was relatively small and limited to just 1 year. Assessing STARD criteria per our systematic, replicative, and consultative methodology is labor-intensive. We believe that this first assessment of reporting quality of articles on diagnostic accuracy studies will provide an important benchmark for any related research in the future.

No comparable study has similarly assessed reporting quality of diagnostic accuracy studies in urology. But related studies in other specialties have been published. In 2015, a study in radiology was published by Korevaar et al. that analyzed 112 articles on diagnostic accuracy studies in 12 high-impact journals in 2012; Korevaar et al. assessed compliance with the 2003 STARD criteria, comparing findings to the publication years of 2004 and 2000 [14]. Overall, they found that the mean number of 2003 STARD criteria met was 15.3 ± 3.9 on a scale of 0–25 (with higher scores reflecting higher-quality reporting). That result corresponded to a reporting compliance rate of 61%, which was similar to our reporting compliance rate of 63% per the updated 2015 STARD criteria, on a scale of 0–30 instead. Korevaar et al. noted an improvement of 1.7 points (95% confidence interval [CI] 0.9–2.5) in 2012, as compared with 2004, and an improvement of 3.4 points (95% CI 2.6–4.3), as compared with 2000.

In 2017, a study was published by Gallo that analyzed 23 full-text articles published in eight emergency medicine journals. Using the 2003 STARD checklist, they found that just over half of the criteria were reported in more than 80% of the articles. Adherence to individual criteria ranged from 8.7% to 100% [15]. Also in 2017, a study was published by Toews that analyzed articles published in 231 hematology and oncology journals; Toews assessed to what extent STARD and other reporting guidelines were referred to in the journals' instructions for authors [16]. The rates of formal journal endorsement of STARD criteria were 3.6% in 2010, 4.2% in 2012, and 9.2% in 2015; the corresponding rates for endorsement of CONSORT criteria were much higher for those same years: 25.1%, 25.8%, and 31.9%. Clearly, there appears to be less awareness of STARD criteria, which could account for the suboptimal reporting compliance rate that we found in urologic journals. Toews suggested a positive association between journal impact factor and reporting compliance as witnessed by more STARD criteria being met, thereby providing additional extrinsic motivation to improve reporting.

With the increasing availability of molecular diagnostic tests, as well as anatomic and functional imaging in urology, diagnostic accuracy questions have been pushed to the forefront of clinical decision-making and health policy. For example, MRI has been widely adopted to help diagnose new cases of prostate cancer and to aid in the active surveillance of patients with early-stage prostate cancer. In urologic oncology, positron emission tomography (PET) is increasingly used. As with questions about therapy (which are best addressed through randomized controlled trials, for which there are well-defined methodologic and reporting criteria), questions about diagnostic accuracy would be best addressed through widely agreed-upon criteria. Unfortunately, diagnostic accuracy studies have garnered much less attention. Our study is the first of its kind in urology. Given the increasing impact of diagnostic modalities in the practice of urology—including the potential threats of overdiagnosis, overtreatment, and wasteful use of precious health care resources—standards must be raised to ensure reporting quality and compliance with key criteria [17, 18].

The findings of this study represent a cross-sectional snapshot in time that provides the basis for similar studies in the future. We hope that it will increase awareness among urologists of methodologic criteria governing articles of diagnostic accuracy studies and ultimately lead to higher-quality studies. Specific improvements are needed in reporting study registration, sources of funding, and sample size determination. Whereas prospective registration of randomized clinical trials has become the norm, this is not yet

the case for studies of diagnostic accuracy, although existing registries do allow for it [19]. Journal editors should not only endorse reporting guidelines but also enforce their implementation [20]. Future studies should provide a longitudinal assessment of the reporting quality over the course of several years and include a larger set of journals with broader representation of topics as well as evaluate the quality of systematic reviews of diagnostic accuracy studies; well-done reviews are critical to the development of effective clinical practice guidelines and sound health care and insurance policies [7].

Conclusion

The reporting quality of studies of diagnostic accuracy published in the urological literature appears modest and independent of clinical topic, journal of publication or study size. This study is limited by its focus on a single publication year of four major journals and preponderance of oncology-related studies, in particular prostate cancer. There is an urgent need for greater awareness for the reporting quality of these types of studies among readers, editors and investigators in order to raise evidentiary standards for questions of diagnosis.

Acknowledgement We thank Mary Knaterrud, Ph.D. for her careful editorial review of the grammar and wording of this manuscript.

Author's contribution DWS: protocol/project development, data collection and management, manuscript writing and editing SG: data collecting and management, manuscript editing PD: protocol/project development, data analysis, manuscript writing and editing

Compliance with ethical standards

Conflict of interest Nothing to disclose.

Research involving human participants and/or animals For this type of study formal consent is not required. This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

References

1. Simera I, Altman DG (2009) ACP journal club. editorial: writing a research article that is "fit for purpose": EQUATOR network and reporting guidelines. Ann Intern Med 151:JC2-2–JC2-3
2. Schulz KF, Altman DG, Moher D et al (2010) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. PLoS Med 7:e1000251
3. Narayan VM, Cone EB, Smith D et al (2016) Improved reporting of randomized controlled trials in the urologic literature. Eur Urol 70:1044–1049
4. Bossuyt PM, Reitsma JB, Bruns DE et al (2003) The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Ann Intern Med 138:W1–12
5. Scales CD Jr, Dahm P, Sultan S et al (2008) How to use an article about a diagnostic test. J Urol 180:469–476
6. Bossuyt PM, Reitsma JB, Bruns DE et al (2015) STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. BMJ 351:h5527
7. Han JL, Gandhi S, Bockoven CG et al (2017) The landscape of systematic reviews in urology (1998–2015): an assessment of methodological quality. BJU Int 119:638–649
8. Tseng TY, Breau RH, Fesperman SF et al (2008) Evaluating the evidence: the methodological and reporting quality of comparative observational studies of surgical interventions in urological publications. BJU Int 103:1026–1031
9. Lokker C, Haynes RB, Wilczynski NL et al (2011) Retrieval of diagnostic and treatment studies for clinical use through PubMed and PubMed's Clinical Queries filters. JAMIA 18:652–659
10. Covidence systematic review software, Melbourne, Australia, Veritas Health Innovation. www.covidence.org. Accessed 20 Aug 2018
11. Research registry. www.researchregistry.com. Accessed 20 Aug 2018
12. McGinn T, Wyer PC, Newman TB et al (2004) Tips for learners of evidence-based medicine: 3. measures of observer variability (kappa statistic). CMAJ 171:1369–1373
13. Ahmed HU, El-Shater Bosaily A, Brown LC et al (2017) Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. Lancet 389:815–822
14. Korevaar DA, Wang J, van Enst WA et al (2015) Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. Radiology 274:781–789
15. Gallo L, Hua N, Mercuri M et al (2017) Adherence to standards for reporting diagnostic accuracy in emergency medicine research. Acad Emerg Med 24:914–919
16. Toews I, Binder N, Wolff RF et al (2017) Guidance in author instructions of hematology and oncology journals: a cross sectional and longitudinal study. PLoS One 12:e0176489
17. Esserman LJ, Thompson IM, Reid B (2013) Overdiagnosis and overtreatment in cancer: an opportunity for improvement. JAMA 310:797–798
18. Macleod MR, Michie S, Roberts I et al (2014) Biomedical research: increasing value, reducing waste. Lancet 383:101–104
19. Korevaar DA, Hooft L, Askie LM et al (2017) Facilitating prospective registration of diagnostic accuracy studies: a STARD initiative. Clin Chem 63:1331–1341
20. Kunath F, Grobe HR, Rucker G et al (2012) Do journals publishing in the field of urology endorse reporting guidelines? A survey of author instructions. Urol Int 88:54–59