Clinical Anatomy 00:00–00 (2016)

ORIGINAL COMMUNICATION

Development of the Anatomical Quality Assessment (AQUA) Tool for the Quality Assessment of Anatomical Studies Included in Meta-Analyses and Systematic Reviews

BRANDON MICHAEL HENRY,[1,2]* KRZYSZTOF A. TOMASZEWSKI,[1,2] PIRAVIN KUMAR RAMAKRISHNAN,[1,2] JOYEETA ROY,[1,2] JENS VIKSE,[1,2] MARIOS LOUKAS,[3] R. SHANE TUBBS,[4] AND JERZY A. WALOCHA[1,2]

[1]*International Evidence-Based Anatomy Working Group, Krakow, Poland*
[2]*Department of Anatomy, Jagiellonian University Medical College, Krakow, Poland*
[3]*Department of Anatomical Sciences, St. George's University, Grenada*
[4]*Seattle Science Foundation, Seattle, Washington, USA*

Critical appraisal of anatomical studies is essential before the evidence from them undergoes meta-epidemiological synthesis. However, no instrument for appraising anatomical studies with inherent applicability to different study designs is available. We aim to develop a generic yet comprehensive tool for assessing the quality of anatomical studies using a formal consensus method. The study steering committee formulated an initial conceptual design and generated items for a preliminary tool on the basis of a literature review and expert opinion. A Delphi procedure was then adopted to assess the validity of the preliminary tool. Feedback from the Delphi panelists was used to improve it. The Delphi procedure involved 12 experts in anatomical research. It comprised two rounds, after which unanimous consensus was reached about the items to be included. The preliminary tool consisted of 20 items, which were phrased as signaling questions and organized into five domains: 1. Aim and subject characteristics, 2. Study design, 3. Characterization of methods, 4. Descriptive anatomy, and 5. Results reporting. Each domain was set to end with a risk of bias question. Following round 1, some of the items underwent major revision, although agreement was reached regarding inclusion of all the domains and signaling questions in the preliminary tool. The tool was revised only for minor language inaccuracies after round 2. The AQUA Tool was designed to assess the quality and reliability of anatomical studies. It is currently undergoing a validation process. Clin. Anat. 00:000–000, 2016.

Key words: anatomy; tool; quality assessment; bias; validity

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Brandon Michael Henry, Department of Anatomy, Jagiellonian University Medical College, 12 Kopernika St, 31–034 Krakow, Poland. E-mail: bmhenry55@gmail.com

Collaborators: Nihal Apaydin; Franciszek Burdan; Matthew L Carlson; Stephen Carmichael; Fabrice Duparc; Alicja Kędzia; Martin M. Mortazavi; Konstantinos Natsis; Mohammadali Shoja; Susan Standring; Joel Vilensky; Stephanie Woodley

Received 26 February 2016; Revised 30 September 2016; Accepted 4 October 2016

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/ca.22799

INTRODUCTION

Comprehensive assessment of a study is crucial for reaching an informed decision regarding its reliability and practical implications. In systematic reviews and meta-analyses, critical appraisal of the included studies, especially of their methodological rigor, lays the foundation for reliable presentation of empirical evidence concerning the subject. The importance of thorough assessment of studies prior to evidence and/or recommendation synthesis has been strongly stressed (Jadad et al., 1996; Higgins and Green, 2011). As a result, numerous approaches to study assessment have been designed and proposed to facilitate meta-epidemiological studies (Yammine, 2014; Henry et al., 2016). The field of anatomy is also experiencing a burgeoning of such evidence-based approaches, which aim to advance the frontiers of clinical anatomy. However, as in other areas of research, most published anatomical studies also have distinctive inherent deficiencies (improper anatomical definitions, dubious measurements of outcome-of-interest, poor observation descriptions, etc.). This limits the capacity of meta-epidemiological studies to provide quantitative summaries of anatomical evidence that are valid, reliable, and applicable.

Understanding the differences among terms used to describe the multiple facets of study appraisal is the key to developing a highly reliable and effective assessment tool. "Quality," "bias," and "validity" are three terms often used interchangeably in assessing studies, though none of them is completely descriptive of the others (Hartling et al., 2009). "Quality" is an abstract concept that is highly subjective and difficult to quantify. It is often interpreted as the strength of study design and implementation, and the ability to preclude systematic errors or bias (Centre for Reviews and Dissemination, 2009; Hartling et al., 2009; Higgins and Green, 2011). Quality is also viewed from various perspectives including, but not limited to, appropriateness of study design, risk of bias, choice of outcome measure, statistical issues, reporting quality, intervention quality, and generalizability (Centre for Reviews and Dissemination, 2009). "Bias," a concept conveniently incorporated into quality, can be defined as systematic deviations from the true fundamental effect owing to poor study design or execution in the collection, analysis, interpretation, publication or review of data (Centre for Reviews and Dissemination, 2009). "Validity," in contrast, is judged from two dimensions; internal and external. Internal validity refers to experimental conduct and avoidance of confounding or bias, while external validity indicates generalizability or applicability of the study findings to other settings. It is important to acknowledge that these elements overlap, which demonstrates that a tool exclusively addressing the assessment of "quality," "bias," or "validity" does not reflect the true reliability of the study. The current challenge is to construct a tool that strikes a balanced assessment of "quality," "bias," and "validity" among anatomical studies, with appropriate emphasis on each element.

Statistical pooling of results from studies of low quality (methodological and/or reporting) produces false evidence (Schuit et al., 2015). For instance, meta-analysis of prevalence data from poor studies on anatomical variations of a particular structure would generate inaccurate and non-generalizable findings, i.e., findings not representative of the general population. The International Evidence-Based Anatomy Working Group (iEBA-WG) strongly believes that for anatomical studies, methodological and reporting qualities are equally important in gauging overall reliability and reproducibility, hallmarks of a "high quality" study. It is vital to remember that high reporting quality does not mean that the study also has high methodological quality or low susceptibility to bias (Sanderson et al., 2007). Therefore, we sensed the need for a consensus regarding critical elements that need to be assessed before the quality of an anatomical study can be determined. Consequently, these elements are integrated into an anatomical study quality assessment tool. Ideally, the tool should be concise, reliable, simple and easy to use. We feel it is also imperative to decide between a domain and a summary scoring system for the tool, taking their strengths and weaknesses into consideration. Appreciating the diversity among anatomical studies (gross, microscopic, surface, surgical, radiological, developmental, electrophysiological, etc.), the tool should also be generic yet comprehensive in assessing study design and topic-specific items.

In brief, the purpose of this project is to develop a quality assessment tool for anatomical studies using a formal consensus method.

MATERIALS AND METHODS

The AQUA Tool was conceived and developed in 2015 with the creation of a steering committee comprising the authors of this study. It espoused the process highlighted by Whiting et al. (Whiting et al., 2003) in their article on the development of the QUADAS tool, which followed the approach suggested by Streiner and Norman health measurement scales (Streiner et al., 1995).

The AQUA Tool was developed in four stages: (1) preliminary conceptual design and item generation; (2) face validity assessment; (3) consistency and construct validity using field trials; and finally (4) generation of a refined AQUA Tool.

The steering committee carefully reviewed the literature to probe the quality of reporting in anatomical studies to develop a preliminary conceptual design. The committee then began item generation, which eventually developed into a preliminary tool. A Delphi protocol was then conducted to assess the face validity of the preliminary tool and refine it on the basis of feedback from experienced anatomists around the world. The Delphi procedure employed in our study was devised and modified from that used in the development of the QUADAS tool (Whiting et al., 2003). An overview of the development process is presented in Figure 1.

Preliminary Conceptual Design and Item Generation

Our primary aim was to develop a tool for assessing anatomical study quality that would aid evidence-

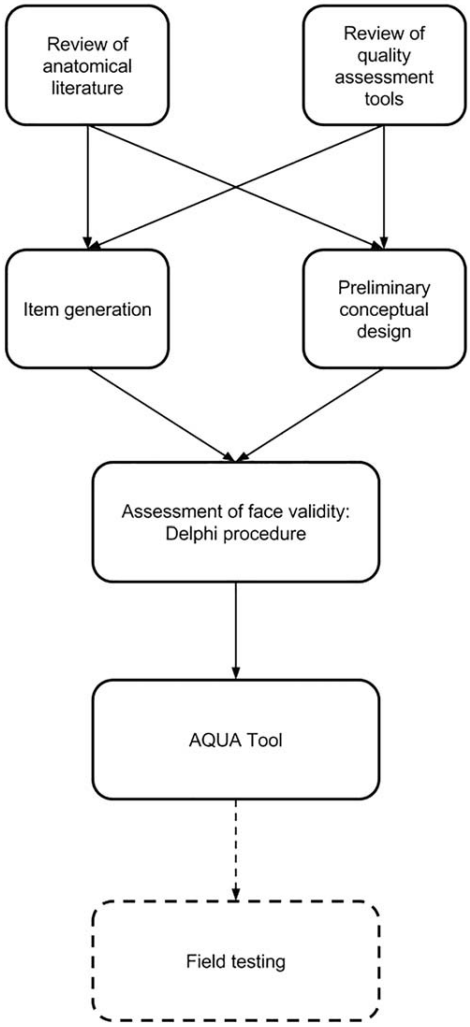

Fig. 1. Overview of the development process of the AQUA tool.

based anatomy (systematic reviews and meta-analyses). Ideally, the tool would be quick, consistent, and reliable, but applicable to a wide range of anatomical study types (microscopic, cadaveric, imaging, intraoperative, etc.). Ease of interpretability would also be an inherent quality. Our aim was to design a tool to answer the primary question: assessing the degree to which the results of a study "should be believed" (Higgins and Green, 2011).

The steering committee unanimously defined "quality" in an anatomical study to include internal validity, strong enough methodological description to allow reproducibility, and clarity and consistency in reporting of both study data and anatomical descriptions.

After reviewing other quality assessment tools for primary studies, the steering committee unanimously agreed to adopt a domain system instead of a scoring system because the latter has less capacity to assess the methodological and reporting quality of a study accurately (Greenland and O'Rourke, 2001). A scoring system also entails greater variability and inconsistency among reviewers in applying weighting to different study domains, which is a major established limitation (Sanderson et al., 2007). The domain-based structure adopted for this tool is similar in design to other tools for assessing bias in original studies, such as QUADAS-2 (Whiting et al., 2011) and the Cochrane Risk of Bias Tool (Higgins and Green, 2011). The steering committee agreed to use signaling questions that assessed an aspect of methodology or reporting, with options of "yes," "no," or "unclear," to facilitate the judgment of risk of bias for each specific domain. It was agreed that risk of bias in each domain would be judged "high," "low," or "unclear."

Items for the AQUA Tool were generated on the basis of (1) the extensive background of the steering committee in conceiving and executing anatomical meta-analyses and systematic reviews, and (2) a review of the anatomical literature (Andall et al., 2015; Henry et al., 2015a, 2015b; Ramakrishnan et al., 2015; Roy et al., 2015; Tomaszewski et al., 2016a, 2016b, 2016c, 2016d, 2016e, 2016f; Vikse et al., 2016). No comprehensive systematic review of systematic reviews was performed because there was no previously-used quality assessment of anatomical studies, and the number of true evidence-based reviews and meta-analyses in the literature is limited. In addition to aspects of methodological quality, the steering committee agreed to focus on quality of reporting and important aspects of descriptive anatomy in developing the preliminary tool. After a list of preliminary items had been developed, they were organized into domains. A signaling question was then developed for each item, and a risk of bias question was produced for each domain.

Delphi Procedure

The steering committee agreed to assess the face validity of the preliminary AQUA Tool using a Delphi Procedure. Potential panelists for the procedure were selected on the basis of their experience in anatomy or evidence-based research and were invited by email to participate in the study. The panel members included editors-in-chief of major anatomical journals and their editorial boards, committee members of reputable anatomical societies, editors-in-chief of major anatomical textbooks and atlases, and other distinguished experts in the fields of anatomy or evidence-based research methods. The Delphi panel members were invited from all major continents to obtain a global input into the development of the tool. After all responses from each round of the Delphi procedure had been received from all panelists, the steering committee carefully assessed each of them. Following deliberations among the authors, the preliminary tool was revised on the basis of the panelists' recommendations, with unanimous agreement among the members of the steering committee.

Delphi Round 1

The Delphi round 1 judged the face validity and quality of the AQUA Tool and comprised an online

survey. The reviewers were asked to assess each signaling question and domain-specific bias using a 5-point Likert scale (5—Strongly agree, 4—Moderately agree, 3—Neutral, 2—Moderately disagree, 1—Strongly disagree), and to assess domain quality using a second Likert Scale (5—Excellent, 4—Very Good, 3—Good, 2—Fair, 1—Poor). A mean score of >4 for an item indicated the need for minor revision, 2–4 for major revision, and <2 for either major revision or serious reconsideration for inclusion in the tool. At the end of the survey, the reviewers were also encouraged to comment on each item, to suggest edits or provide feedback (strengths and weaknesses), and on further steps by which the overall quality and usefulness of the AQUA Tool could be improved.

Delphi Round 2

In Delphi round 2, a revised version of the AQUA Tool was sent to all the panelists who had participated in the first round, along with a point-by-point response to all comments provided by each panelist in Delphi round 1, with justifications for the revisions. The steering committee asked the panelists to assess the revised version of the tool for any additional edits or comments, and encouraged further feedback.

Consistency and Construct Validity

The validity of the AQUA Tool is currently being assessed in field trials. For this purpose, the tool will be implemented on a small sample of published anatomical studies with particular emphasis on its consistency and reliability.

Generation of a Refined AQUA Tool

The feedback and analysis from the field testing will be used to finalize the AQUA Tool.

RESULTS

Item Generation

A list of 20 items for possible inclusion in the quality assessment tool was produced by the steering committee. These items were phrased as signaling questions (which are answered as "Yes," "No," or "Unclear"), and were subsequently organized into five domains: 1. Aim and subject characteristics, 2. Study design, 3. Characterization of methods, 4. Descriptive anatomy, and 5. Results reporting. Each domain was also set to end with a risk of bias question (judged as "Low," "High," or "Unclear"). Following this, the steering committee unanimously agreed that the preliminary AQUA Tool (Supporting Information 1) was ready for assessment of face validity (the Delphi procedure).

Delphi round 1

A total of 20 experts in the field of anatomical or evidence-based research were invited to take part in the Delphi procedure. Twelve of them agreed to participate in round 1 and completed the online surveys. The overall mean Likert score for all the signaling, domain-specific bias, and domain quality rating questions was 4.12 ± 0.39. Of the 20 signaling questions, four items received a mean Likert score of <4: #2 and #5 of Domain 3 (Characterization of Methods), and #1 and #2 of Domain 4 (Descriptive Anatomy). The expert panelists reached agreement regarding inclusion of all the domains and signaling questions in the preliminary AQUA Tool. However, a number of suggestions and comments were made regarding language corrections and improvements of clarity or rephrasing of the questions. The steering committee considered these suggestions and comments carefully and made relevant amendments where appropriate.

Delphi round 2

All 12 panelists from round 1 participated in Delphi round 2. This round did not include an online survey. The expert panelists approved all the modifications made by the steering committee and provided several minor language corrections to ensure consistency. Following these revisions, the AQUA Tool was once again reviewed by the steering committee, who unanimously agreed the tool was ready for field testing.

User's Guide to the AQUA Tool

The revised version of the tool contains five domains (Table 1, Supporting Information 2), each with a set of signaling questions to help assess and judge the risk of bias pertaining it. The signaling questions are answered as "Yes," "No," or "Unclear," indicating low, high, and unclear risks of bias, respectively. Conversely, the risk of bias question is judged as "Low," "High," or "Unclear." If all signaling questions for a domain are answered "Yes," then risk of bias can be judged "Low." If any signaling question is answered "No," this indicates the potential for bias. The reviewer should then reach a consensus on this point. The "Unclear" option should be used only when the reported data are insufficient to allow for clear judgment. However, we emphasize that if the signaling questions could not be answered owing to unreported or missing information, the risk of bias should be judged as "High."

To reduce subjectivity in the risk of bias assessment, each study should be individually and independently examined for bias by at least two members of the review team. In the event of a disagreement among the reviewers, a decision should be reached on the basis of deliberations and consensus among the entire review team.

Reviewers using the AQUA Tool should take care in assessing a study not to be too rigid with respect to the structure of the article matching the structure of the tool itself. For example, Domain 3—Characterization of Methods—contains a signaling question concerning whether the images presented in the study accurately reflect the methods/techniques used. However, although this is part of the assessment of

TABLE 1. List of Domains with their Signaling Questions and Risk of Bias Judgment as Included in the Revised Version of the AQUA Tool

	Options		
Domains & Questions	Yes	No	Unclear
Domain 1: OBJECTIVE(S) AND SUBJECT CHARACTERISTICS			
Was (Were) the objective(s) of the study clearly defined?			
Was (Were) the chosen subject sample(s) and sample size appropriate for the objective(s) of the study?			
Are the baseline and demographic characteristics of the subjects (age, sex, ethnicity, healthy or diseased, etc.) appropriate and clearly defined?			
Could the method of subject selection have in any way introduced bias into the study?	RISK: LOW/HIGH/UNCLEAR		
Domain 2: STUDY DESIGN			
Does the study design appropriately address the research question(s)?			
Were the materials used in the study appropriate for the given objective(s) of the study?			
Were the methods used in the study appropriate for the given objective(s) of the study?			
Was the study design, including methods/techniques applied in the study, widely accepted or standard in the literature? If "no", are the novel features of the study design clearly described?			
Could the study design have in any way introduced bias into the study?	RISK: LOW/HIGH/UNCLEAR		
Domain 3: METHODOLOGY CHARACTERIZATION			
Are the methods/techniques applied in the study described in enough detail for them to be reproduced?			
Was the specialty and the experience of the individual(s) performing each part of the study (such as cadaveric dissection or image assessment) clearly stated?			
Are all the materials and methods used in the study clearly described, including details of manufacturers, suppliers etc.?			
Were appropriate measures taken to reduce inter- and intra-observer variability?			
Do the images presented in the study indicate an accurate reflection of the methods/techniques (imaging, cadaveric, intraoperative, etc.) applied in the study?			
Could the characterization of methods have in any way introduced bias into the study?	RISK: LOW/HIGH/UNCLEAR		
Domain 4: DESCRIPTIVE ANATOMY			
Were the anatomical definition(s) (normal anatomy, variations, classifications, etc.) clearly and accurately described?			
Were the outcomes and parameters assessed in the study (variation, length, diameter, etc.) appropriate and clearly defined?			
Were the figures (images, illustrations, diagrams, etc.) presented in the study clear and understandable?			
Were any ambiguous anatomical observations (i.e., those likely to be classified as "others") clearly described/depicted?			
Could the description of anatomy have in any way introduced bias into the study?	RISK: LOW/HIGH/UNCLEAR		
Domain 5: REPORTING OF RESULTS			
Was the statistical analysis appropriate?			
Are the reported results as presented in the study clear and comprehensible, and are the reported values consistent throughout the manuscript?			
Do the reported numbers or results always correspond to the number of subjects in the study? If not, do the authors clearly explain the reason(s) for subject exclusion?			
Are all potential confounders reported in the study, and subsequently measured and evaluated, if appropriate?			
Could the reporting of results have in any way introduced bias into the study?	RISK: LOW/HIGH/UNCLEAR		

methods in the tool, such images may be presented in other parts of the manuscript (e.g., results section), and should not be judged to entail a higher risk of bias if not presented in the methods. Demographic and baseline characteristics of the subjects participating in the study can variously be reported in either the methods or the results section, and the placement should not have implications for risk of bias assessment. However, any methods or results presented only in the discussion section of the manuscript should always be viewed with considerable caution.

Following field testing and validation, the AQUA Tool will be made freely available online on the website of the International Evidence-Based Anatomy Working Group (http://www.eba.cm.uj.edu.pl/). To help authors to prepare their risk of bias assessments, Microsoft Office templates for tabular and graphical presentation of quality assessment results will also be made freely available on the website. Lastly, after field testing and refinement, a detailed guide will be presented on the use of the AQUA Tool, including examples and suggestions for incorporating the results of risk bias assessment into systematic reviews and meta-analyses of anatomical studies.

DISCUSSION

The importance of high quality anatomical studies should be emphasized since synthesis of findings from such studies serves as the foundation for translational research and clinical interventions. Quality assessment is as crucial for anatomical studies as for other types of research (clinical trials, observational studies, meta-analyses, systematic reviews, etc.). Nevertheless, anatomical research is unique in its own way, although it is in essence observational. Several elements are highly specific to anatomical studies, such as descriptions of normal and variant anatomy and of ambiguous observations. Critical appraisal of anatomical studies is therefore influenced by their methodological quality and their reporting of results. The AQUA Tool judges the quality of an anatomical study by assessing risks of bias due to the methods and the reporting of results. Hence, readers of an anatomical study (reviewers, clinicians, or other researchers) can to assess its reliability on the basis of its risks of bias, as revealed by the tool. The developers of the tool emphasize that it is not designed to provide an overall assessment of the quality of an anatomical article (which is affected by study motivation, synthesis of the literature, methodological approach, data quality, discussion of findings, conclusions, limitations, etc.). It is also not intended to serve as a replacement for reviewer guidelines in journals.

The AQUA Tool assesses the risks of bias in an anatomical study from the perspective of five key domains before providing information about reliability and quality. There are several reasons for adopting a domain-specific risk of bias assessment system rather than a quality assessment score or scale system. The central problem with quality scores is the weighting of individual component items. The lack of empirical evidence renders it impossible to allocate similar or different weightings to the various items of quality assessment (Sanderson et al., 2007). Secondly, the objective and true summary quality score of a study is difficult to determine owing to the relatively arbitrary and subjective process of choosing and calculating quality scores (Whiting et al., 2003). Moreover, quality scores do not recognize that the importance of individual items and their associated level of potential bias can vary according to context (Greenland, 1994; Jüni et al., 1999; Whiting et al., 2003). Since "quality" in anatomical studies was defined by the steering committee to include internal validity, strength of methodological descriptions, and clarity and consistency of reporting of results, a scale system was not chosen because it has not been shown to provide more reliable assessments of validity (Jüni et al., 1999), and could pose a greater risk of confusing the reporting quality with the study validity (Higgins and Green, 2011). Scales and scores are also less likely to be transparent to readers regarding the multiple aspects of quality assessment, and could conceal important strengths and weaknesses in a study (Shea et al., 2009). Moreover, a scale or score system takes more time to complete than a simple domain-specific risk of bias assessment system (Higgins and Green, 2011).

The development of the AQUA Tool presents many advantages to the initiatives undertaken in evidence-based anatomy (EBA). The tool primarily enables the risks of bias to be assessed in studies included in anatomical meta-analyses, the lack of which is a significant limitation in EBA (Henry et al., 2016). It can also be utilized to determine a threshold for inclusion of studies. Failure of a study to satisfy multiple signaling questions (i.e., quality assessment items) could indicate high risks of bias (across one or more domains), thus providing potential reason(s) for its exclusion. Conversely, it can provide a possible explanation for discrepancies in results among anatomical studies and suggest the need for further statistical analyses (e.g., subgroup and sensitivity). The first recorded attempt to appraise the included anatomical studies critically in a review, performed by Smith et al. is noteworthy (Smith et al., 2008). However, the lack of information regarding the development of their appraisal tool and the assessment of its validity and feasibility are significant limitations. Wilke et al. developed the QUality Appraisal for Cadaveric Studies (QUACS) scale to assess the quality of observational cadaveric studies (Wilke et al., 2015). Although this tool was also developed using an expert consensus process and assessed for reliability, validity, and feasibility, it employs a scale system and is only applicable to observational cadaveric studies. Furthermore, the QUACS scale considers elements that are not directly related to methodological quality and risk of bias, and includes several items regarding the overall quality of an article such as "clinical implication of the results are discussed" and "limitation of the results are addressed." Considering the strengths and weaknesses of the tools available in anatomical research and observational studies (Sanderson et al., 2007; Smith et al., 2008; von Elm et al., 2008; Wilke et al., 2015), the AQUA Tool was designed to be applicable across different

types of anatomical studies (gross, microscopic, surface, surgical, radiological, developmental, electrophysiological, etc.) since they might be pooled together into a single analysis. Its construction has the flexibility to assess design-specific areas of various studies in the field. In addition, it could provide clarity and objectivity during critical appraisal of such studies, and point out improvements in the design, conduct, and reporting of future studies.

The project and the AQUA Tool have several limitations. The tool was developed on the basis of expert consensus without empirical evidence for potential sources of bias in anatomical studies. This is because true anatomical meta-analyses and systematic reviews are scarce, and evidence or information regarding bias and/or variability is little-reported in the field. Next, prior to risk of bias assessment, it is prudent to consider factors such as the peer-review process, editorial policy, or journal space restriction that could cause discrepancies between the methodological and reporting qualities (Sanderson et al., 2007). Reviewers should attempt to obtain additional data by contacting the original study authors when necessary. Because of poor reporting quality, most anatomical studies, especially those dating from previous decades, are likely to have high risks of bias across the assessment domains. This can make it challenging to interpret the study findings, making it difficult to establish inclusion and exclusion criteria. At present, we would discourage the setting of arbitrary cutoffs for excluding studies from an analysis, and recommend caution in deciding what to exclude. Quality assessment, even with the AQUA Tool, is subjective and could demonstrate variable inter-rater and intra-rater reliability. We therefore encourage two or more reviewers to perform quality assessment of the studies independently. Any disagreements during the process should be settled by a consensus among all the authors. Finally, the tool does not assess other potential bias-prone areas not directly related to methods and result reporting in anatomical studies. Examples are the accuracy of the conclusions in the light of the results, description of study limitations, disclosure of conflicts of interest, and so on. However, we feel these elements are strongly associated with the reporting quality, and have addressed them appropriately in the AQUA Checklist for reporting original anatomical research. Contrary to the tool which functions to assess the risk of bias in studies included in anatomical systematic reviews and meta-analyses, the checklist serves as a guideline for authors in reporting original anatomical research.

The AQUA steering committee firmly believes that additional studies with a focus on the reproducibility and construct validity of the AQUA Tool are absolutely imperative before solid recommendations and suggestions on its use can be made. Currently, the tool is undergoing rigorous evaluation. Any problems or weaknesses of the tool highlighted during this process will be taken into consideration for its further improvement. Through field testing, we would like to explore problems relating to quality assessment of poorly-reported anatomical studies, as well as issues of subjectivity of assessment and inter-rater and intra-rater reliabilities. The reliability, construct validity, and feasibility of the tool will also be assessed. Following validation, we will probe statistical methods to include risk of bias assessment in the presentation of results in a meta-analysis, based on the different levels of quality of the included studies. The AQUA steering committee strongly feels that the tool is an evolving instrument and requires continuous appraisal and modifications. We therefore welcome any comments, feedback, or recommendations from the scientific community to improve the tool.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

Andall RG, Matusz P, du Plessis M, Ward R, Tubbs RS, Loukas M. 2015. The clinical anatomy of cystic artery variations: A review of over 9800 cases. Surg Radiol Anat 38:529–539.

Centre for Reviews and Dissemination. 2009. Systematic reviews: CRD's guidance for undertaking reviews in health care. York: University of York.

Greenland S. 1994. Quality scores are useless and potentially misleading: reply to "Re: A critical look at some popular analytic methods.". Am J Epidemiol 140:300–301.

Greenland S, O'Rourke K. 2001. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics 2:463–471.

Hartling L, Ospina M, Liang Y, Dryden DM, Hooton N, Krebs Seida J, Klassen TP. 2009. Risk of bias versus quality assessment of randomised controlled trials: Cross sectional study. BMJ 339:b4012.

Henry BM, Roy J, Ramakrishnan PK, Vikse J, Tomaszewski KA, Walocha JA. 2015a. Association of migraine headaches with anatomical variations of the Circle of Willis: Evidence from a meta-analysis. Neurol Neurochir Pol 49:272–277.

Henry BM, Zwinczewska H, Roy J, Vikse J, Ramakrishnan PK, Walocha JA, Tomaszewski KA. 2015b. The prevalence of anatomical variations of the median nerve in the carpal tunnel: A systematic review and meta-analysis. PLoS One 10:e0136477.

Henry BM, Tomaszewski KA, Walocha JA. 2016. Methods of evidence-based anatomy: A guide to conducting systematic reviews and meta-analysis of anatomical studies. Ann Anat - Anat Anzeiger 205:16–21.

Higgins JPT, Green S, 2011. Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration.

Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, McQuay HJ. 1996. Assessing the quality of reports of randomized clinical trials: Is blinding necessary?. Control Clin Trials 17:1–12.

Jüni P, Witschi A, Bloch R, Egger M. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 282:1054–1060.

Ramakrishnan PK, Henry BM, Vikse J, Roy J, Saganiak K, Mizia E, Tomaszewski KA. 2015. Anatomical variations of the formation and course of the sural nerve: A systematic review and meta-analysis. Ann Anat 202:36–44.

Roy J, Henry BM, Pękala PA, Vikse J, Saganiak K, Walocha JA, Tomaszewski KA. 2015. Median and ulnar nerve anastomoses in the upper limb: A meta-analysis. Muscle Nerve 54:36–47.

Sanderson S, Tatt ID, Higgins JPT. 2007. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. Int J Epidemiol 36:666–676.

Schuit E, Roes KCB, Mol BWJ, Kwee A, Moons KGM, Groenwold RHH. 2015. Meta-analyses triggered by previous (false-)significant findings: problems and solutions. Syst Rev 4:57.

Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M. 2009. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 62:1013–1020.

Smith TO, Nichols R, Gilding E, Donell ST. 2008. Are location, proportion and length of VM patellar attachment aetiological factors in patellofemoral dysfunction? A systematic review. Eur J Orthop Surg Traumatol 19:63–73.

Streiner DL, Norman GR, Cairney J. 1995. Health measurement scales: A practical guide to their development and use. Oxford University Press: Oxford, United Kingdom.

Tomaszewski KA, Graves MJ, Henry BM, Popieluszko P, Roy J, Pękala PA, Hsieh WC, Vikse J, Walocha JA. 2016a. The surgical anatomy of the sciatic nerve: A meta-analysis. J Orthop Res 34:1820–1827.

Tomaszewski KA, Henry BM, Vikse J, Pękala PA, Roy J, Svensen M, Guay D, Hsieh WC, Loukas M, Walocha JA. 2016b. Variations in the origin of the deep femoral artery: A Meta-Analysis. Clin Anat (New York, N.Y.).

Tomaszewski KA, Henry BM, Vikse J, Roy J, Pękala PA, Svensen M, Guay DL, Saganiak K, Walocha JA. 2016c. The origin of the medial circumflex femoral artery: A meta-analysis and proposal of a new classification system. PeerJ 4:e1726.

Tomaszewski KA, Popieluszko P, Graves MJ, Pękala PA, Henry BM, Roy J, Hsieh WC, Walocha JA. 2016d. The evidence-based surgical anatomy of the popliteal artery and the variations in its branching patterns. J Vasc Surg.

Tomaszewski KA, Popieluszko P, Henry BM, Roy J, Sanna B, Kijek MR, Walocha JA. 2016e. The surgical anatomy of the lateral femoral cutaneous nerve in the inguinal region: A meta-analysis. Hernia 20:649–657.

Tomaszewski KA, Roy J, Vikse J, Pękala PA, Kopacz P, Henry BM. 2016f. Prevalence of the accessory deep peroneal nerve: A cadaveric study and meta-analysis. Clin Neurol Neurosurg 144: 105–111.

Vikse J, Henry BM, Roy J, Ramakrishnan PK, Hsieh WC, Walocha JA, Tomaszewski KA. 2016. Anatomical variations in the sinoatrial nodal artery: A meta-analysis and clinical considerations. PLoS One 11:e0148331.

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. 2008. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. J Clin Epidemiol 61:344–349.

Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijnen J. 2003. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25.

Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MMG, Sterne JAC, Bossuyt PMM. 2011. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 155:529–536.

Wilke J, Krause F, Niederer D, Engeroff T, Nürnberger F, Vogt L, Banzer W. 2015. Appraising the methodological quality of cadaveric studies: validation of the QUACS scale. J Anat 226: 440–446.

Yammine K. 2014. Evidence-based anatomy. Clin Anat 27:847–852.