

Development of the Documentation and Appraisal Review Tool for systematic reviews

Rebecca L Diekemper, Belinda K Ireland, Liana R Merz

Rebecca L Diekemper, American College of Chest Physicians, CHEST, Glenview, IL 60026, United States
Belinda K Ireland, The EvidenceDoc, Pacific, MO 63069, United States
Liana R Merz, Center for Clinical Excellence, BJC HealthCare, Saint Louis, MO 63108, United States

Revised: April 2, 2015
Accepted: April 27, 2015
Article in press: April 29, 2015
Published online: June 26, 2015

Author contributions: Diekemper RL was the primary developer of the tool and she participated in the testing of the tool and drafting parts of the paper; Ireland BK came up with the concept of developing the tool and was a co-developer of the tool and participated in the testing of the tool and drafting parts of the paper; Merz LR was a co-developer of the tool and participated in the testing of the tool and drafting parts of the paper.

Conflict-of-interest: All of the authors report that they receive no financial compensation for DART. Diekemper RL uses DART for assessing the quality of systematic reviews used to inform guideline recommendations for CHEST guidelines. Due to her role as a developer of DART, the tool has been adopted by CHEST for use in guideline development. Ireland BK reports that as a consultant who frequently conducts systematic reviews and overviews of reviews, she is interested in an effective and efficient tool for evaluating the quality of systematic reviews. Merz LR has no conflicts of interest to disclose.

Data sharing: No additional data are available.

Open-Access: This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

Correspondence to: Rebecca L Diekemper, MPH, American College of Chest Physicians, 2595 Patriot Blvd, Glenview, IL 60026, United States. rdiekemper@chestnet.org
Telephone: +1-314-5319325

Received: January 27, 2015
Peer-review started: February 5, 2015
First decision: March 6, 2015

Abstract

AIM: To develop a tool to more explicitly assess and document the quality of systematic reviews.

METHODS: We developed the Documentation and Appraisal Review Tool (DART) using epidemiologic principles of study design and the following resources: the modified Overview Quality Assessment Questionnaire (modified OQAQ), Assessment of Multiple Systematic Reviews (AMSTAR), the Cochrane Handbook, and the standards promoted by the Agency for Healthcare Research and Quality, and the Institutes of Medicine (IOM). We designed the DART tool to include the following: more detail to provide guidance and improve standardization of use, an approach to assess quality of systematic reviews addressing a variety of research designs, and additional space for recording notes to facilitate recall. DART underwent multiple rounds of testing with methodologists of varying levels of training and experience. Based on the results of six phases of pilot testing, we revised DART to improve performance, clarity and consistency. Pilot testing also included comparisons between DART, and the two most commonly used tools to evaluate the quality of systematic reviews, the modified OQAQ and AMSTAR.

RESULTS: Compared to AMSTAR and modified OQAQ, DART includes two unique questions and several questions covered by modified OQAQ or AMSTAR but not both. Modified OQAQ and DART had the highest reporting consistency. Four AMSTAR questions were unclear and elicited inconsistent responses. Identifying reviewer rationale was most difficult using the modified OQAQ tool, and easiest using DART. DART allows

for documentation of reviewer rationale, facilitating reconciliation between reviewers and documentation for future updates. DART also provides a comprehensive, systematic approach for reviewers with limited experience with systematic review methodology, to critically analyze systematic reviews. In addition, DART is the only one of the three tools to explicitly include quality review for biases specific to observational studies. This is now more widely recognized as important for assessing risk in order to generate recommendations that balance benefit to harm. The tool also includes the assessment of standards recommended by the March 2011 IOM Standards for Systematic Review.

CONCLUSION: This comprehensive tool improves upon existing tools for assessing the quality of systematic reviews and guides reviewers through critically analyzing a systematic review.

Key words: Quality assessment tool; Methodology; Healthcare research; Systematic review; Meta-analysis; Guidelines

© **The Author(s) 2015.** Published by Baishideng Publishing Group Inc. All rights reserved.

Core tip: Systematic reviews and meta-analyses are commonly used to inform the recommendations presented in evidence-based clinical practice guidelines. The purpose of this study was to evaluate the Documentation and Appraisal Review Tool (DART) for its comprehensiveness, identify areas addressed by DART that were not addressed by two other validated tools [Overview Quality Assessment Questionnaire (OQAQ) and Assessment of Multiple Systematic Reviews (AMSTAR)], and to test its performance in eliciting consistent responses. We found that our tool was more comprehensive and included several questions not included in the other tools. We also found that DART elicited the most consistent responses when compared to OQAQ and AMSTAR.

Diekemper RL, Ireland BK, Merz LR. Development of the Documentation and Appraisal Review Tool for systematic reviews. *World J Meta-Anal* 2015; 3(3): 142-150 Available from: URL: <http://www.wjgnet.com/2308-3840/full/v3/i3/142.htm> DOI: <http://dx.doi.org/10.13105/wjma.v3.i3.142>

INTRODUCTION

Systematically collected and critically evaluated evidence forms the backbone of evidence-based clinical practice guidelines, hospital order sets, and quality measurement. Grant *et al*^[1] define a systematic review as a systematic search, appraisal and synthesis of research evidence, often adhering to guidelines for conducting a review. Systematic reviews are the most comprehensive and valid method of collecting and synthesizing the published and unpublished record of clinical science, making

them a preferred source of evidence and encouraging increased production. In 2010, Bastian *et al*^[2] estimated 11 systematic reviews are published each day.

The consistent application of well-defined processes is essential to creating valid systematic reviews. These processes include (1) development of specific clinical question(s) using an analytic framework and standard format to articulate the question(s); (2) use of comprehensive and systematic methods to search for evidence; (3) unbiased process for selecting relevant research; (4) critical evaluation of the quality of included studies; (5) the extraction and synthesis of data from the included studies; and (6) the use of a pre-specified system to evaluate the body of evidence^[3]. Even though these processes for sound systematic review are well described, and reporting checklists like Preferred Reporting Items for systematic reviews and meta-analyses^[4] are available to authors to ensure a higher quality systematic review, the quality of published systematic reviews is not uniformly high. In 2002, Shea *et al*^[5] evaluated the quality of Cochrane and other systematic reviews published in paper based journals, using the Oxman and Guyatt scale and the Sacks checklist. They found the average quality low for both types of reviews.

The Institute of Medicine (IOM) recognized that variation in the quality of systematic reviews still exists and convened a panel in 2010 to develop national standards for the design and implementation of systematic reviews. In 2011, the IOM panel released a list of 21 recommended standards for conducting systematic reviews^[3]. If implemented properly and consistently, these standards could greatly reduce the variability and improve the overall quality of systematic reviews.

Currently, providers and policy makers wanting to incorporate the findings from existing systematic reviews into care decisions, protocols, and guidelines need assistance in evaluating the quality of systematic reviews. Several tools have been developed and evaluated and two have been validated for content^[5,6]. We reviewed published user experience with these two, the modified Overview Quality Assessment Questionnaire (modified OQAQ)^[5] and the Assessment of Multiple Systematic Reviews (AMSTAR)^[6]. Most current users report implementation of AMSTAR because methods for evaluating systematic reviews have advanced since the development of OQAQ, however some also report modifying AMSTAR because it did not meet all their needs^[7,8]. The Agency for Healthcare Research and Quality (AHRQ) recommends that its Evidence-based Practice Centers (EPCs) supplement the use of AMSTAR with additional considerations when incorporating existing systematic reviews into their reviews^[8].

We examined both tools for use in evaluating systematic reviews of clinical interventions in a health system setting. Neither met all our needs (Table 1), and so we first set out to enhance one of the existing assessment tools. However, ultimately we determined the need to develop a comprehensive tool that improves

Table 1 Assessment of existing systematic review quality assessment tools

Need	Modified OQAQ	AMSTAR
Standardized quality assessment process across multiple reviewers with varying levels of experience	Insufficient detail to evaluate disputes	Confusing questions leading to inconsistent responses by same reviewer as well as between reviewers
Single tool to assess a variety of included research designs including randomized trials and observational studies	Insufficient detail on methods	Insufficient detail on methods
Detailed record of the review to facilitate updates of the evidence review	Insufficient detail for replication	Confusing questions leading to inconsistent responses by same reviewer and insufficient detail for replication
Training tool for junior epidemiologists and interns in systematic review methods	Insufficient detail on methods	Insufficient detail on methods

OQAQ: Overview Quality Assessment Questionnaire; AMSTAR: Assessment of Multiple Systematic Reviews.

upon existing tools for assessing the quality of systematic reviews and that guides reviewers through critically analyzing a systematic review. Here we describe the development of a tool designed to more explicitly document the quality assessment of systematic reviews: the Documentation and Appraisal Review Tool (DART) for Systematic Reviews (Table 2). To download the complete tool, please go to <http://www.theevidencedoc.com>.

MATERIALS AND METHODS

Design

DART was developed using epidemiologic principles of study design, the AMSTAR tool^[6], and the Cochrane Handbook for Systematic Reviews of Interventions (version 4.2.6)^[9] as guides. Once completed, we compared our tool to the validated systematic review tools, modified OQAQ and AMSTAR, and to tools developed by some of the AHRQ EPCs to ensure that the tool was as comprehensive as possible. All questions in the DART tool include the following: more detail to provide guidance and improve standardization of use, an approach to assess quality of systematic reviews addressing a variety of research designs, and additional space for recording notes to facilitate recall.

First round testing

An internal group of six methodologists then reviewed and pilot-tested the tool. The group was given systematic reviews of varying quality and asked to use the tool to critically analyze the reviews. The group met weekly for several weeks, testing a different systematic review with the tool each week. This exercise resulted in several revisions. By the end of phase II, we determined that the tool was designed well enough to elicit consistent responses and agreement regarding the overall quality of the studies reviewed.

Comparison of test performance to validated tools

The second round of testing focused on the review of systematic reviews using DART in addition to the modified OQAQ and AMSTAR, two widely accepted, validated tools for assessing the quality of systematic reviews. The goal of this round of testing was to compare

the performance of DART to the modified OQAQ and AMSTAR to determine if we met our design goals. Four internal reviewers with varying levels of training and experience, ranging from a student enrolled in a Masters of Public Health program to a faculty epidemiologist with over 30 years of experience used the three tools to independently assess the quality of several published systematic reviews. The reviewers then used a modified nominal group technique to brainstorm the strengths, weaknesses, and suggestions for improvement of DART. The reviewers also compared the performance of the three tools and identified variation in the responses to the quality assessment questions. The three tools were then mapped against each other to identify and characterize areas of overlap between the questions (Table 3), in order to determine if design goals for DART were met.

Refinement

After evaluating results from the content mapping and comparing performance and utility of DART for reviewers with different levels of experience, the tool was once again revised. A third round of pilot testing was performed using the revised tool to appraise the quality of different systematic reviews.

Comparison to IOM standards for systematic reviews

As a final review of our tool, we compared content to the March 2011 Standards for Systematic Reviews from the IOM to ensure that the tool included an evaluation component for each IOM standard^[3].

Final testing

Final modification of the tool was completed in April 2011, followed by more rounds of internal pilot testing to evaluate consistency of responses for each question when the same reviewer appraised the systematic review at different points in time (intra-observer reliability) and when used by different reviewers (inter-observer reliability).

RESULTS

Assessing comparability of content of the three tools

In order to determine if we met our design goals, we

Table 2 Documentation and Appraisal Review Tool for systematic reviews

Title of Systematic Review:			
Author:			
Publication date:		Article tracking number:	
Reviewer:		Date completed:	
1 Did the authors develop the research question(s) and inclusion/exclusion criteria before conducting the review?			Use this space to document the rationale for your answer
a	It was clear the authors developed the research question(s) and inclusion criteria before conducting the review and that they stated the question(s) clearly	Yes	
b	Not described or cannot tell	No	
2 Did the authors describe the search methods used to find evidence (original research) on the primary question(s)?			Use this space to document the rationale for your answer
a	Key words and/or MESH terms were stated and where feasible the search strategy was provided	Yes	
b	Not described or cannot tell	No	
3 Was the search for the evidence reasonably comprehensive? Were the following included?			Use this space to document the rationale for your answer
a	Search included at least two electronic sources	Yes	No
b	Authors chose the most applicable electronic databases (<i>e.g.</i> , CINAHL for nursing journals, EMBASE for pharmaceutical journals, and MEDLINE for general, comprehensive search) and only limited search by date when performing an update of a previous systematic review	Yes	No
c	Search methods are likely to capture all relevant studies (<i>e.g.</i> , includes languages other than English; gray literature such as conference proceedings, dissertations, theses, clinical trials registries and other reports) and authors hand-searched journals or reference lists to identify published studies which were not electronically available	Yes	No
4 Did the authors do the following when selecting studies for the review?			Use this space to document the rationale for your answer
a	Provide in the inclusion criteria: population, intervention, outcome and study design?	Yes	No
b	State whether the selection criteria were applied independently by more than one person?	Yes	No
c	State how disagreements were resolved during study selection?	Yes	No
d	Provide a flowchart or descriptive summary of the included and excluded studies?	Yes	No
e	Include all study designs appropriate for the research questions posed?	Yes	No
5 Were the characteristics of the included studies provided? (in an aggregated form such as a table, data from the original studies were provided on the participants, interventions and outcomes)			Use this space to document the rationale for your answer
a	Yes		
b	Partially		
c	No		
6 Did the authors make any statements about assessing for publication bias?			Use this space to document the rationale for your answer
a	The authors did assess for publication bias and if publication bias was detected they stated how it was handled	Yes	
b	The authors did assess for publication bias but did not state how it was handled if it was detected	Partially	
c	Not described or cannot tell	No	
7 Did the authors do the following to assess the overall quality of the individual studies included in the review?			Use this space to document the rationale for your answer
a	Was the quality assessment specified with adequate detail to permit replication?	Yes	No
b	Was the quality assessment conducted independently by more than one person?	Yes	No
c	Did the authors state how disagreements were resolved during the quality assessment?	Yes	No
8 Did the authors appropriately assess for quality by appropriately examining the following sources of bias in all of the included studies?			Use this space to document the rationale for your answer
All studies:			
a	Confounding (assessed comparability of study groups at start of study, was randomization successful?)	Yes	No
b	Sufficient sample size (only applicable to studies that summarize their results in a qualitative manner; it's not a concern for pooled results)	Yes	No
c	Outcome reporting bias (assessed for each outcome reported using a system such as the ORBIT classification system)	Yes	No
d	Follow up (assessed for completeness and any differential loss to follow-up)	Yes	No
For Randomized Controlled Trials only:			
e	Randomization	Yes	No
f	Allocation concealment	Yes	No
g	Blinding	Yes	No

For Case-Control and Cohort Studies only:			
h	Selection bias	Yes	No
i	Information bias--recall and completeness to follow-up	Yes	No
For Quasi-Experimental Studies only:			
j	Differences between the first and second study measurement point - such as changes or improvements in other interventions, changes in measurement techniques or definitions, or aging of subjects	Yes	No
k	Selection bias	Yes	No
For Diagnostic Accuracy Studies only:			
l	Selection (spectrum) bias - were subjects selected to be representative of patients to whom the test will be applied in clinical practice, and to represent the broadest spectrum of disease?	Yes	No
m	Verification bias - were all patients subjected to the same reference standard of diagnosis, and was it measured blindly and independently of the test?	Yes	No
9 Did the authors use appropriate methods to extract data from the included studies?		Use this space to document the rationale for your answer	
a	Were standard forms developed and piloted prior to the systematic review conduct?	Yes	No
b	Did the authors ensure that data from the same study but that appeared in multiple publications were counted only once in the synthesis?	Yes	No
c	Was data extraction performed by more than one person?	Yes	No
10 Did the authors assess and account for heterogeneity (differences in participants, interventions, outcomes, trial design, quality or treatment effects) among the studies selected for the review?		Use this space to document the rationale for your answer	
a	The authors stated the differences among the studies and how they accounted for those differences	Yes	
b	The authors stated the differences but not how they accounted for them	Partially	
c	Not described or cannot tell	No	
11 Did the authors describe the methods they used to combine/synthesize the results of the relevant studies (to reach a conclusion) and were the methods used appropriate for the review question(s)?		Use this space to document the rationale for your answer	
a	Methods were reported clearly enough to allow for replication. The overview included some assessment of the qualitative and quantitative heterogeneity of the study results and the results were appropriately combined/synthesized. For meta-analyses, an accepted pooling method (<i>i.e.</i> , more than simple addition) was used. Or the authors state that the evidence is conflicting and that they can't combine/synthesize the results	Yes	
b	The methods were reported clearly enough to allow for replication but they were not combined appropriately	Partially	
c	Not described or cannot tell	No	
12 Did the authors perform sensitivity analyses on any changes in protocol, assumptions, and study selection? (For example, using sensitivity analysis to compare results from fixed effects and random effects models)		Use this space to document the rationale for your answer	
a	Sensitivity analyses were used when appropriate on all changes in a priori design	Yes	
b	Sensitivity analyses were only used on some changes in a priori design	Partially	
c	Not described or cannot tell	No	
13 Are the conclusions of the authors supported by the reported data with consideration of the overall quality of that data?		Use this space to document the rationale for your answer	
a	The conclusions are supported by the reported data and reflect both the scientific quality of the studies and the risk of bias in the data obtained from those studies	Yes	
b	The authors failed to consider study quality and/or their conclusions were not supported by the data, or cannot tell	No	
14 Were conflicts of interest stated and were individuals excluded from the review if they reported substantial financial and intellectual COIs?		Use this space to document the rationale for your answer	
a	COIs were reported for each team member and individuals were excluded if they had substantial COIs	Yes	
b	COIs were reported but it was not clear whether individuals were excluded based on their COIs	Partially	
c	COIs were not reported and individuals were not excluded based on their COIs	No	
15 On a scale of 1-10, how would you judge the overall quality of the paper?			
Rating Overall Comments			
Good (8-10)			
Fair (5-7)			
Poor (< 5)			

COIs: Conflicts of interests.

Table 3 Comparison of Documentation and Appraisal Review Tool to modified Overview Quality Assessment Questionnaire and Assessment of Multiple Systematic Reviews

DART questions	Corresponding AMSTAR question(s)	Corresponding modified OQAQ question(s)
(1) Did the authors develop the research question(s) and inclusion/exclusion criteria before conducting the review?	(1) Was an "a priori" design provided?	Not addressed
(2) Did the authors describe the search methods used to find evidence (original research) on the primary question(s)?	(3) Was a comprehensive literature search performed?	(1) Were the search methods used to find evidence on the primary question stated?
(2a) Are key words and/or MESH terms stated?	(3) Was a comprehensive literature search performed?	Not addressed
(3) Was the search for the evidence reasonably comprehensive?	(3) Was a comprehensive literature search performed?	(2) Was the search for evidence reasonably comprehensive?
(3a) Does the search include at least 2 databases?	(3) Was a comprehensive literature search performed?	Not addressed
(3b) Did the authors choose the most applicable electronic databases and only limit the search by date when performing an update?	Not addressed	Not addressed
(3c) Are search methods likely to capture all relevant studies and did the authors hand-search journals or reference lists to identify published studies which were not electronically available?	(3) Was a comprehensive literature search performed?	Not addressed
	(4) Was the status of publication (<i>i.e.</i> , grey literature) used as an inclusion criterion?	
(4a) Did the authors provide in the inclusion criteria: Population, intervention, outcome, and study design, when selecting studies for the review?	Not addressed	Not addressed
(4b) Did the authors state whether the selection criteria were applied by more than one person? ¹	(2) Was there duplicate study selection and data extraction? ¹	Not addressed
(4c) Did the authors state how disagreements were resolved during study selection? ¹	(2) Was there duplicate study selection and data extraction? ¹	Not addressed
(4d) Did the authors provide a flowchart or descriptive summary of the included and excluded studies?	(5) Was a list of studies (included and excluded) provided?	Not addressed
(4e) Did the authors include all study designs appropriate for the research questions posed?	Not addressed	Not addressed
(5) Were the characteristics of the included studies provided? (in an aggregated form such as a table, data from the original studies were provided on the participants, interventions and outcomes)	(6) Were the characteristics of the included studies provided?	Not addressed
(6) Did the authors make any statements about assessing for publication bias?	(10) Was the likelihood of publication bias assessed?	Not addressed
(7a) Was the quality assessment specified with adequate detail to permit replication?	(7) Was the scientific quality of the included studies assessed and documented?	(5) Were the criteria used for assessing the validity of the included studies reported?
(7b) Was the quality assessment conducted independently by more than one person?	Not addressed	Not addressed
(7c) Did the authors state how disagreements were resolved during the quality assessment?	Not addressed	Not addressed
(8) Did the authors appropriately assess for quality by appropriately examining the following sources of bias in all of the included studies: confounding, sufficient sample size, outcome reporting bias, follow-up, randomization, allocation concealment, blinding, selection bias, information bias, verification bias, and differences between the first and second study measurement point?	(7) Was the scientific quality of the included studies assessed and documented? (partial match)	(6) Was the validity of all studies referred to in the text assessed using appropriate criteria? (partial match)
(9) Did the authors use appropriate methods to extract data from the included studies?	Not addressed	Not addressed
(9a) Were standard forms developed and piloted prior to the systematic review conduct?	Not addressed	Not addressed
(9b) Did the authors ensure that data from the same study that appeared in multiple publications were counted only once in the synthesis?	Not addressed	Not addressed
(9c) Was data extraction performed by more than one person?	(2) Was there duplicate study selection and data extraction?	Not addressed
(10) Did the authors assess and account for heterogeneity (differences in participants, interventions, outcomes, and trial design, quality or treatment effects) among the studies selected for the review?	(9) Were the methods used to combine the findings of studies appropriate?	(7) Were the methods used to combine the findings of the relevant studies reported?

(11) Did the authors describe the methods they used to combine/synthesize the results of the relevant studies (to reach a conclusion) and were the methods used appropriate for the review question(s)?	(9) Were the methods used to combine the findings of studies appropriate?	(8) Were the findings of the relevant studies combined appropriately?
(12) Did the authors perform sensitivity analyses on any changes in protocol, assumptions, and study selection? (For example, using sensitivity analysis to compare results from fixed effects and random effects models)	Not addressed	(8) Were the findings of the relevant studies combined appropriately? Not addressed
(13) Are the conclusions of the authors supported by the reported data with consideration of the overall quality of that data?	(8) Was the scientific quality of the included studies used appropriately in formulating conclusions? (partial match)	(9) Were the conclusions made by the author(s) supported by the data reported? (partial match)
(14) Were conflicts of interest stated and were individuals excluded from the review if they reported substantial financial and intellectual COIs?	(11) Was the conflict of interest stated? (partial match)	Not addressed
(15) On a scale of 1-10, how would you judge the overall quality of the paper?	Not addressed	(10) Overall quality

¹Separate questions in DART, but concepts not separated in AMSTAR. DART: Documentation and Appraisal Review Tool; OQAQ: Overview Quality Assessment Questionnaire; AMSTAR: Assessment of Multiple Systematic Reviews; COIs: Conflicts of interests.

mapped OQAQ and AMSTAR to DART and displayed the results in Table 3. Table 3 shows that our tool includes several questions that are unique and not included in the modified OQAQ or AMSTAR, with several other questions covered by one or the other but not both tools.

Assessing consistency of performance of the three tools

Throughout the iterations of development, testing and group discussion and review of performance, we learned that the modified OQAQ and DART consistently produced similar overall assessments of quality. However, during these discussions we had more difficulty remembering or locating reviewer rationale for the responses using the modified OQAQ tool. DART has sufficient space to record page and line details to facilitate recall. This was important when resolving disputes. We also discovered that the AMSTAR tool had questions that were confusing and difficult to implement consistently. They are the following: (1) Question 4: Was the status of publication (*i.e.*, grey literature) used as an inclusion criterion? The authors should state that they searched for reports regardless of their publication type. The authors should state whether or not they excluded any reports from the systematic review, based on their publication status, language, *etc.* This question was confusing since it seemed to equate an accurate description of the extent of the search with the actual execution of a thorough search; (2) Question 5: Was a list of studies (included and excluded) provided? This question was interpreted as being too specific by requiring lists, and did not allow for a good flow chart; it seemed to require more detail than most journal space would allow; (3) Question 7: Was the scientific quality of the included studies assessed and documented? A priori methods of assessment should be provided [*e.g.*, for effectiveness studies if the author(s) chose to include only randomized,

double-blind, placebo controlled studies, or allocation concealment as inclusion criteria]; for other types of studies alternative items will be relevant. This question did not provide sufficient detail to execute consistently. We found it more useful to specify the most important sources of bias by study type for consistent reporting both within and across reviewers; and (4) Question 11: Was the conflict of interest stated? Potential sources of support should be clearly acknowledged in both the systematic review and the included studies. The answer to this question was always no. Systematic review authors often mention their personal sources of support, but we did not find an example where potential sources of support were provided for the included studies. This needs to either be two questions, or allow for partial scoring.

DART was the only one of the three tools to explicitly include quality review for biases specific to observational studies. Since the importance of including evidence from observational data is now more widely recognized, particularly for assessing risk in order to generate recommendations that balance benefit to harm, we believe it is important to include careful assessment of the potential for biased measurement unique to this design.

DISCUSSION

We are aware that a revision of the AMSTAR tool exists and is known as R-AMSTAR^[7]. The primary goal for revising AMSTAR was to produce an overall quantitative estimate of the quality of the systematic review. The performance of R-AMSTAR has been compared to the original tool using systematic reviews from the field of assisted reproduction for subfertility^[10]. In that comparison study, R-AMSTAR was noted to provide more guidance to the reviewer than AMSTAR, but was

more difficult to apply consistently. Popovich *et al.*^[10] reported that the R-AMSTAR criteria were difficult to apply because of subjectivity of some of the domains, especially domain 8. That question “Was the scientific quality of the included studies used appropriately in formulating conclusion?” provided four criteria, which Popovich *et al.*^[10] report as being difficult to distinguish. Their kappa statistics also showed poor inter-rater reliability for this domain.

We designed the DART quality assessment tool to address limitations we discovered when using the modified OQAQ and AMSTAR tools. The specific improvements are: (1) Space for enhanced recording detail to facilitate reconciliation between reviewers and provide detailed reference for use in future updates; (2) An evaluation of major biases relevant to observational study designs and the assessment of standards recommended by the March 2011 IOM Standards for Systematic Review^[3]; (3) Additional detail and guidance for junior epidemiologists, clinicians and other members of the review panel with less experience in systematic review methods; and (4) Consistent overall quality assessment of systematic reviews using a qualitative ranking that categorizes studies as good, fair or poor at the end of a detailed assessment.

In order to facilitate the use of systematic reviews, the American College of Chest Physicians (CHEST) adopted DART to assess the quality of systematic reviews included in their evidence reviews. CHEST guideline authors used DART to assess the quality of systematic reviews and meta-analyses included in the “Diagnosis and Management of Lung Cancer: CHEST Evidence-Based Clinical Practice Guideline (3rd Edition)”^[11], and subsequent guidelines. DART has been used for other CHEST guidelines and it is discussed in the article Methodologies for the Development of CHEST Guidelines and Expert Panel Reports^[12].

This paper describes the development of DART for systematic reviews. The next step is to quantify the performance of components of the tool through validation testing, assessing inter-rater agreement scores. Based on our preliminary evaluation with the modified OQAQ and AMSTAR, intra-rater reliability should also be tested when assessing the same systematic review at a later point in time, since updated evidence reviews are essential to ensuring that the best current evidence informs clinical guidelines and policy. The ability to facilitate accurate recall of prior reviews will improve the efficiency of that process.

The authors now have considerable experience and familiarity with DART and can complete the assessment form quickly. It is therefore important to use an external validation process to test performance in persons with a wide variety of backgrounds and without prior experience with the tool in order to evaluate inter and intra rater consistencies in response and time for completion.

Well-executed systematic reviews now form the foundation of evidence-based clinical practice guidelines. Even though the IOM has developed rigorous standards

for conducting systematic reviews, there is still wide variation in how they are conducted and reported. Given this variation and the new reliance on systematic reviews, comprehensive tools are needed to assess the quality of systematic reviews. By creating the DART for Systematic Reviews we attempted to fill this gap.

ACKNOWLEDGMENTS

We would like to thank the interns in the Center for Clinical Excellence at BJC HealthCare for assisting us with testing DART and giving us feedback on modifications to the tool.

COMMENTS

Background

Systematic reviews are the foundation for evidence-based guidelines. Rigorous standards exist, but there is wide variation in implementation, highlighting the need for a more comprehensive quality assessment tool for systematic reviews.

Research frontiers

As the publication of systematic reviews increases, variability in the quality still exists. Users of systematic reviews need a way to assess the quality of systematic reviews that includes all relevant study designs. Since the importance of including evidence from observational data is now more widely recognized, especially to assess potential for harm, a single tool is needed that includes careful assessment of the potential for biased measurement unique to this design as well as for randomized trials.

Innovations and breakthroughs

The authors designed the the Documentation and Appraisal Review Tool (DART) quality assessment tool to address limitations they discovered when using the modified Overview Quality Assessment Questionnaire and Assessment of Multiple Systematic Reviews tools. The specific improvements include: the ability to record rationale for each criteria; criteria for assessing observational studies and for assessing standards recommended by the Institute of Medicine in 2011; additional guidance to assist less experienced reviewers in assessing the quality of systematic reviews; and consistent overall quality assessment of systematic reviews using a qualitative ranking.

Applications

DART provides a comprehensive, systematic approach for reviewers with limited experience with systematic review methodology, to critically analyze systematic reviews. It also provides a complete record of judgments and decisions made during the assessment to assist reconciliation between reviewers during the current review and for use in future updates.

Terminology

The terminology used in this article reflects the vocabulary familiar to an audience using systematic reviews for decision-making.

Peer-review

The peer reviewers did not report having any concerns about the paper. Reviewer comments included the following: Systematic reviews are the foundation for evidence-based guidelines and are increasing. The article discusses the development of a comprehensive tool that improves upon existing tools for assessing the quality of systematic reviews and that guides reviewers through critically analyzing a systematic review. It has significance to appraise a systematic review.

REFERENCES

- 1 **Grant MJ**, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. *Health Info Libr J* 2009; **26**: 91-108 [PMID: 19490148]
- 2 **Bastian H**, Glasziou P, Chalmers I. Seventy-five trials and eleven systematic reviews a day: how will we ever keep up? *PLoS Med* 2010; **7**: e1000326 [PMID: 20877712]
- 3 **Institute of Medicine (US) Committee on Standards for**

- Systematic Reviews of Comparative Effectiveness Research.** Finding what works in health care: standards for systematic reviews. Eden J, Levit L, Berg A, Morton S, editors. Washington (DC): National Academies Press (US), 2011 [PMID: 24983062]
- 4 **Moher D**, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009; **151**: 264-269, W64 [PMID: 19622511 DOI: 10.7326/0003-4819-151-4-200908180-00135]
 - 5 **Shea B**, Moher D, Graham I, Pham B, Tugwell P. A comparison of the quality of Cochrane reviews and systematic reviews published in paper-based journals. *Eval Health Prof* 2002; **25**: 116-129 [PMID: 11868441]
 - 6 **Shea BJ**, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007; **7**: 10 [PMID: 17302989]
 - 7 **Kung J**, Chiappelli F, Cajulis OO, Avezova R, Kossan G, Chew L, Maida CA. From systematic reviews to clinical recommendations for evidence-based health care: validation of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *Open Dent J* 2010; **4**: 84-91 [PMID: 21088686]
 - 8 **White CM**, Ip S, McPheeters M, Carey TS, Chou R, Lohr KN, Robinson K, McDonald K, Whitlock E. Using existing systematic reviews to replace de novo processes in conducting comparative effectiveness reviews methods guide for effectiveness and comparative effectiveness reviews. Rockville (MD): Agency for Healthcare Research and Quality (US), 2008 [PMID: 21433402]
 - 9 **Higgins JPT**, Green S, editors. Assessment of study quality. cochrane handbook for systematic reviews of interventions 4.2.6 [updated September 2006]. In: The Cochrane Library. Chichester, UK: John Wiley and Sons, Ltd, 2006: 384
 - 10 **Popovich I**, Windsor B, Jordan V, Showell M, Shea B, Farquhar CM. Methodological quality of systematic reviews in subfertility: a comparison of two different approaches. *PLoS One* 2012; **7**: e50403 [PMID: 23300526 DOI: 10.1371/journal.pone.0050403]
 - 11 **Lewis SZ**, Diekemper R, Addrizzo-Harris DJ. Methodology for development of guidelines for lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest* 2013; **143**: 41S-50S [PMID: 23649432]
 - 12 **Lewis SZ**, Diekemper R, Ornelas J, Casey KR. Methodologies for the development of CHEST guidelines and expert panel reports. *Chest* 2014; **146**: 182-192 [PMID: 25010961 DOI: 10.1378/chest.14-0824]

P- Reviewer: Cid J, Gao C, Tang Y, Trohman RG **S- Editor:** Tian YL
L- Editor: A **E- Editor:** Liu SQ





Published by **Baishideng Publishing Group Inc**

8226 Regency Drive, Pleasanton, CA 94588, USA

Telephone: +1-925-223-8242

Fax: +1-925-223-8243

E-mail: bpgoffice@wjgnet.com

Help Desk: <http://www.wjgnet.com/esps/helpdesk.aspx>

<http://www.wjgnet.com>

