Review of Quality Assessment Tools for the Evaluation of Pharmacoepidemiologic Safety Studies

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001362
Article Type:	Research
Date Submitted by the Author:	08-May-2012
Complete List of Authors:	Neyarapally, George; FDA, FDA/CDER/OSE Hammad, Tarek; FDA/CDER/OSE, Epidemiology Pinheiro, Simone; FDA/CDER/OSE, Epidemiology Iyasu, Solomon; FDA/CDER/OSE, Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Qualitative research, Public health
Keywords:	Adverse events < THERAPEUTICS, EPIDEMIOLOGY, THERAPEUTICS

Title: Review of Quality Assessment Tools for the Evaluation of Pharmacoepidemiologic Studies

Running Head: Evaluation of Pharmacoepidemiologic Safety Studies

Authors: George A. Neyarapally, Tarek A. Hammad, Simone P. Pinheiro, Solomon Iyasu

Author Affiliations: Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD

Disclaimer: The views expressed in this article do not necessarily represent those of the Food and Drug Administration.

Acknowledgments: The authors would like to thank Gerald Dal Pan and Judy Staffa for their critical evaluation of the manuscript and thoughtful suggestions.

Funding: The authors did not receive a specific grant or funding for this research.

Ethics: As this study involved a review of existing assessment tools, a formal ethics review was not required.

Corresponding contact author's information:

Dr. George A. Neyarapally
Food and Drug Administration
10903 New Hampshire Avenue
White Oak Bldg. #22, Room 4461
Silver Spring, Maryland 20993
Phone: (301) 796-1134

Word Count: (excluding abstract, table, figures, references): 2731

Key Words: Quality assessment; pharmacoepidemiologic studies; drug safety

Key Points:

- In the context of regulatory safety-related decision making, quality assessment (i.e., assessment of the risk of bias), informs the evaluation of available evidence and enhances the appropriate utilization of available evidence in assessing the balance between benefits and risks of drugs
- The development of a consolidated reporting and quality assessment tool would enhance the consistent, transparent and objective evaluation of pharmacoepidemiologic safety studies. If a tool is developed, it is important to determine if there is a need for tools tailored for specific study designs or if one tool that consolidates these considerations might be helpful.
- Key findings from our review of quality assessment tools include:
    - Many available quality assessment tools do not include critical assessment elements that are specifically relevant to pharmacoepidemiologic safety studies;
    - Most tools do not distinguish between reporting elements and quality assessment attributes;
    - There is a lack of reported considerations on the relative weights to assign to different domains and elements with respect to assessing the quality of these studies.

Summary of Conflict of Interest Statements: None of the authors report any conflicts of interest.

STRUCTURED ABSTRACT

Objectives: Pharmacoepidemiologic safety studies are an important hypothesis-testing tool in the evaluation of drug safety issues in the postmarket period. Despite the potential to produce robust value-added data, the interpretation of findings from such studies can sometimes be challenging because of their well-recognized methodological limitations. Therefore, assessment of the quality of individual studies is essential to evaluating their credibility. The authors critically evaluated the suitability and relevance of available tools for the quality assessment of these studies.

Design: To examine the utility of individual quality assessment tools for the evaluation of observational pharmacoepidemiologic safety studies, we created an a priori assessment framework consisting of domains that include reporting elements (RE) and quality assessment attributes (QAA). Our comprehensive literature search identified distinct assessment tools and the percentage of tools assessing the pre-specified elements and attributes was tabulated.

Results: After identifying 96 distinct assessment tools, we reviewed 61 tools. Most of the tools reviewed were not designed to evaluate pharmacoepidemiologic safety studies. More than 50% of the reviewed tools considered reporting elements under the research aims, analytical approach, outcome definition and ascertainment, study population, and exposure definition and ascertainment domains. Reporting elements under the discussion and interpretation, results and study team domains were considered in less than 40% of the tools. Except for the data source domain, quality attributes were considered in less than 50% of the tools.

Conclusions: Critical issues and research gaps identified include: (1) many tools do not include critical assessment elements relevant to observational pharmacoepidemiologic safety studies; (2) most tools do not distinguish between reporting elements and quality assessment attributes; and (3) a lack of considerations on the relative weights of different domains and elements. The development of a quality assessment tool would facilitate consistent, objective, and evidence-based assessments of pharmacoepidemiologic safety studies.

INTRODUCTION

Several sources of evidence on drug safety issues inform FDA postmarketing safety-related regulatory decisions, including spontaneous case reports, registries, observational pharmacoepidemiologic studies, randomized controlled trials (RCTs), meta-analyses, and other sources. Despite the well known strengths of RCTs in the assessment of drug efficacy, specific issues related to the design, methodology and transparency of experimental studies may limit their ability to fully characterize the safety profile of drugs after marketing approval.[1-4]

Pharmacoepidemiologic studies, typically observational in nature, represent an important hypothesis-testing tool in the evaluation of drug safety issues suspected at the time of approval and for new signals emerging in the postmarket period. In contrast to RCTs, such studies, which typically employ broader inclusion and fewer exclusion criteria and leverage claims or electronic medical record data, might better reflect the real life experience of patients. Furthermore, pharmacoepidemiologic studies afford the ability to investigate rare drug-related adverse effects, examine risks in patient subpopulations, and assess long-term adverse events. Recent health-related legislation will increase the availability and adoption of electronic healthcare data for such studies.[5,6]

Despite the potential of pharmacoepidemiologic safety studies to produce robust value-added data, the interpretation of findings from such studies can sometimes be challenging because of their well-recognized methodological limitations, including various sources of bias and confounding.[7] These limitations also apply to the increasing number of comparative effectiveness epidemiologic studies.[5;8-10] Therefore, assessment of the quality of individual studies is essential to evaluating their credibility. Transparency in reporting on the design, conduct, analysis and results of these studies is a prerequisite for the assessment of the quality of

the evidence; it is first necessary to understand the relevant aspects of the study design, conduct, and analysis, along with the underlying assumptions and rationale behind the key scientific decisions undertaken by the study team to adequately evaluate the credibility of the study.[11]

The FDA recently published a draft guidance on the design, conduct, and reporting of pharmacoepidemiologic safety studies using electronic healthcare data that is designed both to enhance the transparency of reporting of such studies and to encourage critical scientific thinking regarding their design and conduct.[12] In the future, this guidance may improve the credibility of submitted pharmacoepidemiologic studies by shedding light on the pertinent aspects of studies needed to inform the evaluation of the internal and external validity of their findings. However, even if the bar for transparency and reporting of these studies is raised, it will still be necessary to evaluate the contribution of these studies to the available evidence on an emerging drug safety issue. The Grading of Recommendations Assessment, Development and Evaluation evidence assessment framework, used by clinical guideline developers, appropriately separates the initial processes of quality assessment and the weighing of evidence in the formulation of guideline recommendations.[13,14] In the context of regulatory decision making concerning safety issues, the use of quality assessment tools to assess of the risk of bias may add a measure of objectivity to the scientific judgment of the available evidence and improve the quality of decision making. The benefits may not only extend to improving decision making by regulators but also by journal editors and researchers as well as potentially improving the quality of performed studies by stimulating consideration of key aspects of these studies during the development of the study approach and protocol.

Although many tools for the assessment of epidemiologic studies exist,[15-17] most are not specifically designed to evaluate pharmacoepidemiologic safety studies. Recent articles have

suggested the need to develop tools for assessing the quality of these studies.[18-21] A recent publication found that systematic reviewers and meta-analysts are misusing reporting tools like STROBE due to the dearth of validated assessment tools.[22]

The main objective of this article is to critically evaluate the suitability and relevance of available tools for the assessment of pharmacoepidemiologic safety studies. The ultimate goal is to stimulate discussion in the scientific community about the need for specific tools to facilitate the transparent, objective and consistent evaluation of study quality to inform safety-related regulatory decision making.

METHODS

For purposes of this paper, quality assessment tools are defined as qualitative checklists and/or quantitative scales designed to facilitate assessment of the quality of epidemiologic studies.

A priori quality assessment framework

To examine the utility of individual quality assessment tools for the evaluation of pharmacoepidemiologic safety studies, we created an a priori assessment framework, consisting of domains that include reporting elements (RE), and quality assessment attributes (QAA) (Table 1). Based on the expert opinion of senior FDA epidemiologists, concepts drawn from the FDA draft guidance on such studies, and key findings from seminal reviews and tools,[12;23-27] we established the domains pertaining to the design, conduct, and analysis of pharmacoepidemiologic safety studies. Within each domain we listed critical elements that need to be considered for assessing the validity and interpretation of findings from such studies. We made a distinction between the reporting elements (RE) and quality assessment attributes (QAA) for each domain. The selected elements and attributes presented in this Table are not intended to

represent an all-inclusive list of factors, but rather to represent critical aspects impacting the internal and external validity of pharmacoepidemiologic safety studies. Of note, although the quality assessment attributes necessarily involve some subjectivity, their inclusion in an assessment tool would facilitate the consistent and objective consideration and evaluation of key quality attributes across individual studies. As the GRADE framework developers have emphasized, although quality assessment is fundamentally subjective,[28] developing a transparent, consistent approach to assessment of quality is important, especially in the regulatory and clinical arena as patients, healthcare professionals, and sponsors benefit from consistent and transparent assessment of available evidence for use in decision making.

Literature search

A comprehensive literature search of quality assessment tools and reviews of such tools was performed in MEDLINE, EMBASE, and Web of Science. Search terms included: "assessment," "tools," "quality," "medical research," "evidence based research," "evidence based medicine," "meta-analysis," "randomized controlled trials," "biological product," "drug," "pharmaceutical preparation," "biological therapy," "bias," and "epidemiology." A total of 54 references were retrieved from this search. Two independent reviewers identified potentially relevant abstracts (n=26) from the initial literature review (inter-rater reliability > 0.85). Inclusion criteria included quality assessment tools or reviews of such tools developed to evaluate RCTs, observational studies, or meta-analyses. Exclusion criteria consisted of clinical assessment tools, general articles or guidance on quality assessment, and tools or reviews focused strictly on reporting and not addressing quality assessment. After reviewing each article, 10 relevant tools and 3 seminal reviews of tools were identified. Thirteen tools were excluded based on the exclusion criteria above.

The most recent, relevant review articles of tools and some individual tools for assessing the quality of epidemiologic studies were identified.[24-26] The 2007 Sanderson review,[24] the most recent, comprehensive review of tools for assessing quality of epidemiologic studies, served as the starting point for identification of tools.

We also performed Google Scholar searches to identify relevant tools that might not be captured in the aforementioned search strategies. Google searches based on the first 50 hits included the following terms: "tool quality bias"; "quality assessment"; "pharmacoepidemiology"; "quality assessment epidemiology"; "tool quality assessment study"; and "scale quality assessment observational studies." Furthermore, we identified and included an EMA methodological checklist[23] because, although it is a reporting tool, it includes domains and considerations designed to inform safety evaluations made at a drug regulatory agency.

The assessment tools identified were reviewed by investigators based on a priori assessment criteria shown in Table 1. The percentage of tools assessing the pre-specified elements and attributes within domains was tabulated. During our review of each tool, we documented which tools employed some method of tool validation.

RESULTS

Overall, out of 104 tools identified, a total of 96 distinct assessment tools, including 82 tools from the Sanderson review, 6 tools from the initial literature review, 7 from the Google search, and 1 regulatory tool (EMA tool) were considered for review (Figure 1). Out of these, 61 tools were selected for the in-depth review.[13;23;29-31;31;31-74;75-86] Tools exclusively focused on randomized controlled trials, tools focused on clinical assessments, and tools that did not include an explicit assessment framework were excluded (n=35).

More than half of the reviewed tools considered reporting elements for domains including research aims, analytical approach, outcome definition and ascertainment, study population, and exposure definition and ascertainment. Domains related to the discussion and interpretation, results and study team were considered in less than 40% of the reviewed tools. With the exception of the study population/data sources domain, QAA research domains were considered in less than half of the assessment tools, with less than 10% of the tools considering results and study team domains. Many tools did not address all pertinent domains.

Most reviewed tools were not specifically designed to assess epidemiologic studies of drug-related harms. Although the EMA framework was designed to increase transparency of pharmacoepidemiologic studies, it focuses on reporting versus quality assessment. Our review constitutes the only recent comprehensive review of available assessment tools to determine if any are appropriate and sufficient for the evaluation of pharmacoepidemiologic studies. Only a small number of the reviewed tools employed some method of tool validation.[29-35] Most of the tools did not differentiate between reporting elements (RE) and quality assessment attributes (QAA) whereas others stratified by these aspects. A small number of tools included distinct assessment criteria for different epidemiologic study designs (e.g. case-control, cohort). Tools focused more on RE than QAA. Figure 1 displays the percentage of tools that included criteria on the assessment domains and elements.

The proportion of reviewed tools that included reporting elements (RE) and quality assessment attributes (QAA) according to each a-priori defined domain within the framework is shown in Figure 1.

DISCUSSION

Based on the review of currently available tools, there is no specific tool that is adequately designed for the robust evaluation of pharmacoepidemiologic studies of drug safety. No single tool considered all the selected domains and elements and most tools failed to address critical evaluation elements. Making a distinction between RE and QAA is important as even if an element of the study is mentioned in the final report, one must then determine if this was appropriate for the specific study in the context of the drug safety question. Some of the tools specifically made this distinction as we did in our a priori framework. Additionally, important RE and QAA were lacking in most of the tools we reviewed which highlights the need for a tool focused on the evaluation of epidemiologic studies designed to evaluate drug-related harms; this need has been previously identified by others.[87]

Quality attributes related to exposure definition and ascertainment were considered in less than half of the assessment tools, with less than 15% of the tools including RE and QAA pertaining to the comparator group, despite the fact that the selection of a comparator is critical for drug safety and effectiveness trials and epidemiologic studies, as the choice of suboptimal comparators can provide misleading results.[88;89] Only 30% of the tools included quality assessment elements pertaining to the validity and appropriateness of the operational aspects of exposure ascertainment, and only 36% of tools addressed quality attributes of validation of outcome ascertainment approaches. These are important facets of pharmacoepidemiologic safety studies their misclassification may lead to false negative findings regarding the association between a drug and adverse event.

Only about 40% of the tools included QAA on approaches to handle confounding and biases. As observational studies are not randomized, the approaches to handle confounding and

bias are of paramount importance.[7;90;91] This is an important limitation of most tools because there are often uncertainties regarding results from pharmacoepidemiologic studies due to the limitations of electronic healthcare data and complex nature of the practice of medicine.[91] Only a small percentage of tools (28% RE; 18% QAA) included elements on the consideration of study findings in the context of the design, conduct, limitations, and statistical power despite the fact that these elements are essential in assessing implications of study findings.

Some of the tools we reviewed were designed as "all-purpose" assessment tools for evaluation of clinical trials and observational studies, while others focused on a particular study design (e.g., case-control, cohort). It may be useful to create one consolidated, validated tool for evaluating observational pharmacoepidemiologic safety studies focused on general reporting and quality attributes; secondary focused tools for the specific study designs, i.e., case-control and cohort studies, may also be useful due to some of the unique aspects of these designs. By creating such a tool, regulatory agencies, clinical guideline developers, and clinicians could consistently evaluate studies and for decision making. The creation of this tool could be led by an independent expert or academic group, perhaps with input from regulatory agencies.

Although we did not address weighing of importance of different domains and elements based on their relative impact on study contribution to the available streams of evidence, this may be an important consideration in the formulation of an assessment tool. Also, it is not clear if numerical scores are helpful in assessing the quality of epidemiologic studies, as when numerical scores were used to evaluate systematic reviews or meta-analyses of such studies, they did not produce valid results.[92] The appropriate tradeoff between the utility of a tool for review and the comprehensiveness of the evaluation elements has yet to be determined. This is complicated by the lack of validation of most of the available tools. Before these issues can be

addressed, it is first necessary to engage in a broader discussion of the utility of such assessment tools in the evaluation of pharmacoepidemiologic safety studies. It is worth noting that critical assessment elements of pharmacoepidemiologic studies focused on effectiveness may be different than those focused on safety; however, pharmacoepidemiologic comparative effectiveness studies focus on both comparative safety and benefits associated with drugs and thus such elements are not mutually exclusive.[93] Thus, it is important to consider the potential to leverage current efforts to create a validated assessment tool (GRACE checklist[94]) for observational comparative effectiveness pharmacoepidemiologic studies.

Our review has some limitations. The purpose and scope of the tools we reviewed varied greatly. Although we conducted a comprehensive review, there may be tools that we were unable to access or that were published after our search. If a reporting or quality assessment element contained some aspects of the element, we counted this as full representation, even if not all the important sub-elements were included. Each tool was reviewed by one study team member; repeating the evaluations via a second reviewer was deemed unnecessary at this stage as the primary goal of the review was to obtain a broad understanding of the utility of available assessment tools in evaluating pharmacoepidemiologic safety studies based on a preliminary assessment framework. Some factors that may be increasingly relevant in future studies, such as electronic health records with linkages to other data sources like outpatient claims, health information exchanges, or personal health records, were not included in our framework but may be included in a future validated tool. Guidelines and checklists published after the time period of our review have included some elements that may be important for future linked studies which may leverage the increasing availability of these data sources.[95]

In the evaluation of many emerging safety issues, pharmacoepidemiologic safety studies are discussed and may influence safety-related decision making. However, often the quality-driven contribution of such studies is not discussed in a consistent way. The development of an assessment tool based on expert input may facilitate consistent, evidence-based quality assessment of such studies and the subsequent determination of their value based on evaluating the impact of bias on the robustness of a study results, and the interpretation of its findings, within the context of the specific drug safety issue. The framework we developed may serve as a foundation for future development of such a tool. Efforts to improve the evaluation of the contribution of pharmacoepidemiologic safety studies would be consistent with the Agency's focus on strengthening regulatory safety science.[96] If after further consideration and discussions with stakeholders development of a tool to evaluate epidemiologic data for drug safety is pursued, it would be necessary to first determine the scope of the assessment tool as well as steps for its comprehensive validation. Importantly, such a tool would be intended to complement, and not replace, expert clinical, methodological, and statistical expertise necessary to complete a robust evaluation and determination of the contribution of a specific pharmacoepidemiologic safety study to the available evidence for regulatory decision making.

Table 1: Reporting elements and quality assessment attributes according to selected domains and percent representation among reviewed tools

Reporting Elements (REs):	Percent representation	Quality Assessment Attributes (QAAs):	Percent representation
A. Research aims	69%		34%
RE 1: Description of study objectives, research aims, design, study population and data source, exposure, and outcome	69%	QAA 1: Appropriateness of pre-specified aims, design, population, exposure, and outcome to address research aim	34%
B. Study population: data sources	84%		57%
RE 1: Description of participation rates and discontinuation rates	77%	QAA 1: Extent of participation rates and discontinuation rates	56%
RE 2: Description of denominator used for risk assessment	11%	QAA 2: Appropriateness of denominator used for risk assessment	3%
C. Exposure definition and ascertainment	61%		31%
RE 1: Description of operational aspects of exposure ascertainment and definition	49%	QAA 1: Validity and appropriateness of operational aspects used to ascertain and define exposure status	30%
RE 2: Description of blinding of outcome status	21%		
RE 3: Selection of exposure risk window	5%	QAA 2: Appropriateness of selected exposure risk window	3%
RE 4: Description of selected type of users (incident v. prevalent users)	0%	QAA 3: Appropriateness of selected numerator for risk assessment	0%
RE 5: Description of comparison group	10%	QAA 4: Appropriateness of comparison group	5%
D. Safety outcome definition and ascertainment	69%		36%
RE 1: Description of operational aspects of outcome ascertainment and definition	51%	QAA 1: Appropriateness/validity of outcome ascertainment strategies and outcome definition	33%
RE 2: Description of blinding of exposure status from those ascertaining/validating outcomes	25%		
RE 3: Description of follow up time	16%	QAA 2: Adequacy of follow up time to address research question	13%
RE 4: Description of composite outcome, if relevant	0%	QAA 3: Adequacy of composite safety outcome, if relevant	0%
E. Analytical approach	85%		49%
RE 1: Description of analytic approach, including approaches to handle confounding and biases	80%	QAA 1: Appropriateness of described analytic approach	26%
RE 2: Description of a priori sample size/power calculations	44%	QAA 2: Appropriateness of approaches to handle confounding and biases	39%

RE 3: Description of data integration methods, when relevant	3%	QAA 3: Description of a priori sample size/power calculations	21%
RE 4: Description of measures of frequency and association	7%	QAA 4: Appropriateness of data integration methods, when relevant	0%
RE 5: Description of a priori specifications of subgroup analyses	5%		
F. Results	36%		7%
RE 1: Description of main results (unadjusted and adjusted estimates and confidence intervals) and sensitivity analyses	25%	QAA 1: Consistency of primary, secondary, and sensitivity analyses and consistency of confounding effects with known associations	2%
RE 2: Description of patient disposition	15%	QAA 2: Impact of patient disposition on study integrity and generalizability of findings	6%
RE 3: Description of characteristics of population by comparison group	18%		
G. Discussion and interpretation	36%		20%
RE 1: Description of findings in relation to pertinent issues related to study design, conduct, limitations, and power	28%	QAA 1: Consideration of findings in relation to pertinent issues related to study design, conduct, limitations, and power	18%
RE 2: Description of plausibility of findings and clinical significance and discussion/exploration of alternative explanations, comparison with other findings	21%	QAA 2: Discussion of plausibility of findings and clinical significance and discussion of alternative explanations, comparison with other findings	11%
H. Study team	7%		3%
RE 1: Description of study team, conflict of interest, funding sources	7%	QAA 1: Relevance of study team credentials and experience to the research area	0
		QAA 2: Independence of study team and funding sources	3%

Figure 1 Percent representation by domain of reporting and quality aspects considered by the assessment tools

INSERT FIGURE ONE HERE

Reference List

(1) Hammad TA, Pinheiro SP, Neyarapally GA. Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clin Trials* 2011;8:559-570.

(2) Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737-1739.

(3) Pitrou I, Boutron I, Ahmad N, Ravaud P. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-1761.

(4) Clark DW, Coulter DM, Besag FM. Randomized controlled trials and assessment of drug safety. *Drug Saf* 2008;31:1057-1061.

(5) American Recovery and Reinvestment Act. Pub.L. 111-5. 2009.

(6) Bruen BK, Ku L, Burke MF, Buntin MB. More than four in five office-based physicians could qualify for federal electronic health record incentives. *Health Aff (Millwood)* 2011;30:472-480.

(7) Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-337.

(8) Dreyer N, Tunis S, Berger M. Why Observational Studies Should Be Among The Tools Used In Comparative Effectiveness Research. *Health Affairs* 2010;29:1818.

(9) Dreyer N. Do We Need Good Practice Principles for Observational Comparative Effectiveness Research? *GRACE Initiative* [serial online] 2009.

(10) Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Stat Med* 2010;29:1963-1976.

(11) von EE, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-349.

(12) FDA. FDA Draft Guidance: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets; 2011.

(13) Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;19;328:1490.

(14) Guyatt GH, Oxman AD, Vist GE et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.

(15) Deeks JJ, Dinnes J, D'Amico R et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-173.

(16) Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666-676.

(17) Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care* 2004;16:9-18.

(18) Cornelius VR, Perrio MJ, Shakir SA, Smith LA. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiol Drug Saf* 2009;18:1223-1231.

(19) Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 2005;142:1090-1099.

(20) Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.:32.

(21) Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol* 2006;59:765-769.

(22) da Costa BR, Cevallos M, Altman DG, Rutjes A, Egger M. Uses and misuses of STROBE statement: bibliographic study. *BMJ Open* 2011;doi:10.1136/bmjopen-2010-000048 .

(23) EMA. Checklist for Methodological Standards for ENCePP Study Protocols. 2010.

(24) Sanderson. Tools for assessing the quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology* 2007;36:666.

(25) Deeks. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003;7.

(26) Lohr K. Rating the strength of scientific evidence: relevance for quality improvement programs. *International Journal for Quality in Health Care* 2004;16:9-18.

(27) von Elm E. STROBE. *Journal of Clinical Epidemiology* 2008;344.

(28) Balshem H, Helfand M, Schunemann HJ et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-406.

(29) Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-384.

(30) Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of oro-facial pain. *J Dent* 2001;29:451-467.

(31) Margetts BM, Thompson RL, Key T et al. Development of a scoring system to judge the scientific quality of information from case-control and cohort studies of nutrition and disease. *Nutr Cancer* 1995;24:231-239.

(32) Rangel SJ, Kelsey J, Colby CE, Anderson J, Moss RL. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. *J Pediatr Surg* 2003;38:390-396.

(33) Genaidy AM, Lemasters GK, Lockey J et al. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. *Ergonomics* 2007;50:920-960.

(34) Santaguida, Raina. McMaster Quality Assessment Scale of Harms (McHarm) for primary studies . 2008.

(35) Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994;272:101-104.

(36) Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. *J Clin Nurs* 1994;3:271-277.

(37) Cameron I, Crotty M, Currie C et al. Geriatric rehabilitation following fractures in older people: a systematic review. *Health Technol Assess* 2000;4:i-111.

(38) Carneiro AV. Critical appraisal of prognostic evidence: practical rules. *Rev Port Cardiol* 2002;21:891-900.

(39) NHS. Critical Appraisal Skills Programme: appraisal tools. Public Health Resource Unit N, editor. 2003.

(40) Center for Occupational and Environmental Health. Critical appraisal. School of Epidemiology and Health Sciences UoM, editor. 2003.

(41) DuRant RH. Checklist for the evaluation of research articles. *J Adolesc Health* 1994;15:4-8.

(42) Elwood M. Forward projection--using critical appraisal in the design of studies. *Int J Epidemiol* 2002;31:1071-1073.

(43) Esdaile JM, Horwitz RI. Observational studies of cause-effect relationships: an analysis of methodologic problems as illustrated by the conflicting data for the role of oral contraceptives in the etiology of rheumatoid arthritis. *J Chronic Dis* 1986;39:841-852.

(44) Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. *Br Med J (Clin Res Ed)* 1986;292:810-812.

(45) Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional. Wales, editor. 2004.

(46) Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA* 1994;271:1615-1619.

(47) Lichtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case-control studies. *J Chronic Dis* 1987;40:893-903.

(48) Federal Focus I. The London Principles for Evaluating Epidemiologic Data in Regulatory Risk Assessment. *Federal Focus* [serial online] 2004.

(49) Margetts BM, Vorster, Venter. Evidence-Based Nutrition - Review of Nutritional Epidemiological Studies. *South African Journal of Clinical Nutrition* 2002;15:68-73.

(50) University of Montreal. Critical Appraisal Worksheet. 2004.

(51) Mulrow CD, Lichtenstein MJ. Blood glucose and diabetic retinopathy: a critical appraisal of new evidence. *J Gen Intern Med* 1986;1:73-77.

(52) Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.:25.

(53) Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developer's handbook. *Scottish Intercollegiate Guidelines Network* [serial online] 2004.

(54) Parker S. Critical appraisal: Guidelines for the critical appraisal of a paper. *Website* [serial online] 2004.

(55) Zaza S, Wright-De Aguero LK, Briss PA et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18:44-74.

(56) Zola P, Volpe T, Castelli G et al. Is the published literature a reliable guide for deciding between alternative treatments for patients with early cervical cancer? *Int J Radiat Oncol Biol Phys* 1989;16:785-797.

(57) Ariens GA, van MW, Bongers PM, Bouter LM, van der Wal G. Physical risk factors for neck pain. *Scand J Work Environ Health* 2000;26:7-19.

(58) Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 2002;288:728-737.

(59) Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain* 1998;77:1-13.

(60) Carson CA, Fine MJ, Smith MA, Weissfeld LA, Huber JT, Kapoor WN. Quality of published reports of the prognosis of community-acquired pneumonia. *J Gen Intern Med* 1994;9:13-19.

(61) Loney PL, Chambers LW, Bennett KJ, Roberts JG, Stratford PW. Critical appraisal of the health research literature: prevalence or incidence of a health problem. *Chronic Dis Can* 1998;19:170-176.

(62) Corrao G, Bagnardi V, Zambon A, Arico S. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. *Addiction* 1999;94:1551-1573.

(63) Garber BG, Hebert PC, Yelle JD, Hodder RV, McGowan J. Adult respiratory distress syndrome: a systemic overview of incidence and risk factors. *Crit Care Med* 1996;24:687-695.

(64) Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med* 1994;121:11-21.

(65) Jabbour M, Osmond MH, Klassen TP. Life support courses: are they effective? *Ann Emerg Med* 1996;28:690-698.

(66) Kreulen CM, Creugers NH, Meijering AC. Meta-analysis of anterior veneer restorations in clinical studies. *J Dent* 1998;26:345-353.

(67) Krogh CL. A checklist system for critical review of medical literature. *Med Educ* 1985;19:392-395.

(68) Littenberg B, Weinstein LP, McCarren M et al. Closed fractures of the tibial shaft. A meta-analysis of three methods of treatment. *J Bone Joint Surg Am* 1998;80:174-183.

(69) Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988;260:652-656.

(70) Manchikanti L, Singh V, Vilims BD, Hansen HC, Schultz DM, Kloth DS. Medial branch neurotomy in management of chronic spinal pain: systematic review of the evidence. *Pain Physician* 2002;5:405-418.

(71) Meijer R, Ihnenfeldt DS, van LJ, Vermeulen M, de Haan RJ. Prognostic factors in the subacute phase after stroke for the future residence after six months to one year. A systematic review of the literature. *Clin Rehabil* 2003;17:512-520.

(72) Nguyen QV, Bezemer PD, Habets L, Prahl-Andersen B. A systematic review of the relationship between overjet size and traumatic dental injuries. *Eur J Orthod* 1999;21:503-515.

(73) Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989;84:815-827.

(74) Stock SR. Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. *Am J Ind Med* 1991;19:87-107.

(75) van der Windt DA, Thomas E, Pope DP et al. Occupational risk factors for shoulder pain: a systematic review. *Occup Environ Med* 2000;57:433-442.

(76) Bollini P, Garcia Rodriguez LA, Perez GS, Walker AM. The impact of research quality and study design on epidemiologic estimates of the effect of nonsteroidal anti-inflammatory drugs on upper gastrointestinal tract disease. *Arch Intern Med* 1992;152:1289-1295.

(77) Ciliska D, Hayward S, Thomas H et al. A systematic overview of the effectiveness of home visiting as a delivery strategy for public health nursing interventions. *Can J Public Health* 1996;87:193-198.

(78) Cowley DE. Prostheses for primary total hip replacement. A critical appraisal of the literature. *Int J Technol Assess Health Care* 1995;11:770-778.

(79) Effective Public Health Practice Project. Quality Assessment Tools for Quantitative Studies. 2003.

(80) Faculty of Medical and Health Sciences. EPIQ (Effective Practice, Informatics, and Quality Improvement). 2004. University of Auckland.

(81) Fowkes FG, Fulton PM. Critical appraisal of published research: introductory guidelines. *BMJ* 1991;302:1136-1140.

(82) Gyorkos TW, Tannenbaum TN, Abrahamowicz M et al. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994;85 Suppl 1:S8-13.:S8-13.

(83) Spitzer WO, Lawrence V, Dales R et al. Links between passive smoking and disease: a best-evidence synthesis. A report of the Working Group on Passive Smoking. *Clin Invest Med* 1990;13:17-42.

(84) Steinberg EP, Eknoyan G, Levin NW et al. Methods used to evaluate the quality of evidence underlying the National Kidney Foundation-Dialysis Outcomes Quality Initiative Clinical Practice Guidelines: description, findings, and implications. *Am J Kidney Dis* 2000;36:1-11.

(85) El BN, Middel B, van Dijk JP, Oosterhof A, Boonstra PW, Reijneveld SA. Are the outcomes of clinical pathways evidence-based? A critical appraisal of clinical pathway evaluation research. *J Eval Clin Pract* 2007;13:920-929.

(86) Salford University. Evaluation Tool for Quantitative Research Studies. 2001.

(87) Chou R, Aronson N, Atkins D et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010;63:502-512.

(88) Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff (Millwood)* 2010;29:1842-1848.

(89) Psaty BM, Weiss NS. NSAID trials and the choice of comparators--questions of public health importance. *N Engl J Med* 2007;356:328-330.

(90) Strom B. *Pharmacoepidemiology*. 4th ed. John Wiley & Sons Ltd, 2005.

(91) Brookhart MA, Sturmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010;48:S114-S120.

(92) Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 2010;25:603-605.

(93) Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *JAMA* 2010;303:985-986.

(94) GRACE. GRACE Checklist. Accessed at: http: www.graceprinciples.org. 2012.

(95) Hall GC, Sauer B, Bourke A, Brown JS, Reynolds MW, Casale RL. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012;21:1-10.

(96) Hamburg MA. Shattuck lecture. Innovation, regulation, and the FDA. *N Engl J Med* 2010;363:2228-2232.

April 25, 2012
Contributorship statement for submission to BMJ Open
RE: Manuscript ID bmjopen-2012-001632

Below is a description of the authors' contributions to the manuscript.

A. George A. Neyarapally

1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

B. Tarek A. Hammad

1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

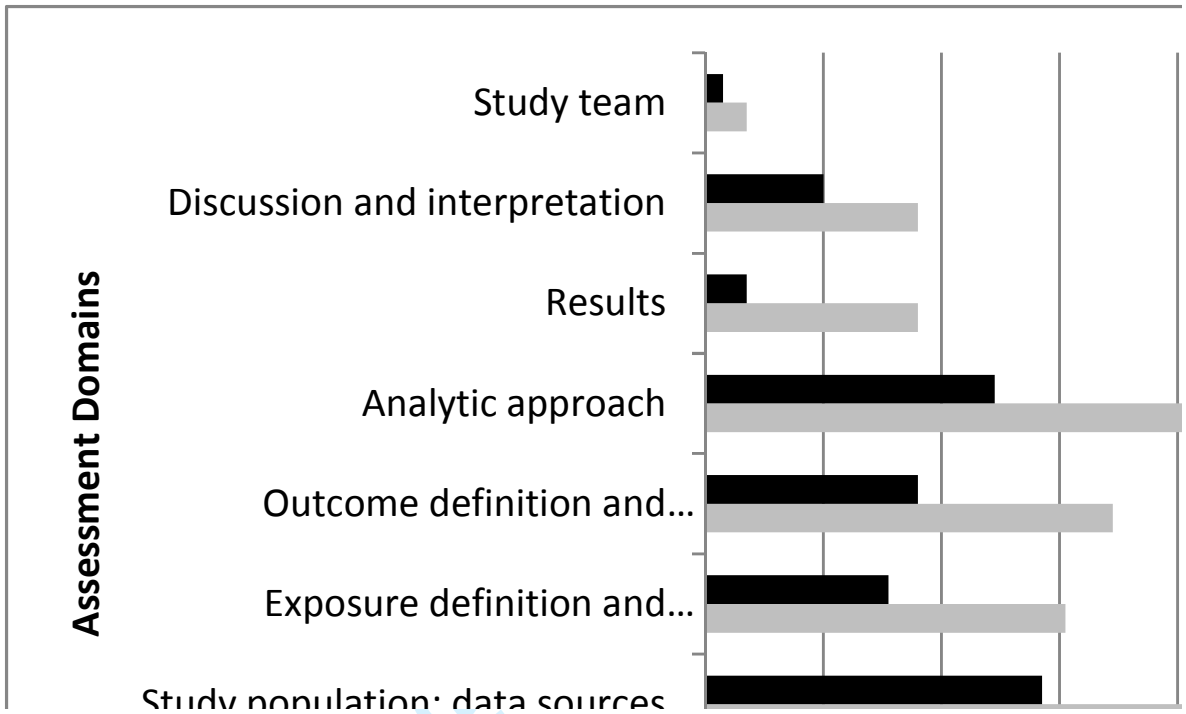
C. Simone P. Pinheiro

1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

D. Solomon Iyasu

1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

Sincerely,
George A. Neyarapally, PharmD, MPH



Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Review of Quality Assessment Tools for the Evaluation of Pharmacoepidemiologic Safety Studies

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2012-001362.R1
Article Type:	Research
Date Submitted by the Author:	26-Jul-2012
Complete List of Authors:	Neyarapally, George; FDA, FDA/CDER/OSE Hammad, Tarek; FDA/CDER/OSE, Epidemiology Pinheiro, Simone; FDA/CDER/OSE, Epidemiology Iyasu, Solomon; FDA/CDER/OSE, Epidemiology
Primary Subject Heading:	Epidemiology
Secondary Subject Heading:	Pharmacology and therapeutics, Public health
Keywords:	Adverse events < THERAPEUTICS, EPIDEMIOLOGY, THERAPEUTICS

7/25/12 Tracked changes revisions to address BMJ Open reviewer comments

Title: Review of Quality Assessment Tools for the Evaluation of Pharmacoepidemiologic Studies
Running Head: Evaluation of Pharmacoepidemiologic Safety Studies
Authors: George A. Neyarapally, Tarek A. Hammad, Simone P. Pinheiro, Solomon Iyasu
Author Affiliations: Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, MD

Disclaimer: The views expressed in this article are those of the authors and do not necessarily represent those of the Food and Drug Administration.
Acknowledgments: The authors would like to thank Gerald Dal Pan and Judy Staffa for their critical evaluation of the manuscript and thoughtful suggestions.
Funding: The authors did not receive a specific grant or funding for this research.
Ethics: As this study involved a review of existing assessment tools, a formal ethics review was not required.

Corresponding contact author's information:
Dr. George A. Neyarapally
Food and Drug Administration
10903 New Hampshire Avenue
White Oak Bldg. #22, Room 4461
Silver Spring, Maryland 20993
Phone: (301) 796-1134

Word Count: (excluding abstract, table, figures, references): 3096
Key Words: Quality assessment; pharmacoepidemiologic studies; drug safety

Key Points:

- In the context of regulatory safety-related decision making, quality assessment (i.e., assessment of the risk of bias), informs the evaluation of available evidence and enhances the appropriate utilization of available evidence in assessing the balance between benefits and risks of drugs
- The development of a consolidated reporting and quality assessment tool would enhance the consistent, transparent and objective evaluation of pharmacoepidemiologic safety studies. If a tool is developed, it is important to determine if there is a need for tools tailored for specific study designs or if one tool that consolidates these considerations might be helpful.
- Key findings from our review of quality assessment tools include:
    - Many available quality assessment tools do not include critical assessment elements that are specifically relevant to pharmacoepidemiologic safety studies;
    - Most tools do not distinguish between reporting elements and quality assessment attributes;

- There is a lack of reported considerations on the relative weights to assign to different domains and elements with respect to assessing the quality of these studies.

Summary of Conflict of Interest Statements: None of the authors report any conflicts of interest.

Funding Statement
This research received no specific funding.

Contributorship Statement
A. George A. Neyarapally
1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

B. Tarek A. Hammad
1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

C. Simone P. Pinheiro
1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

D. Solomon Iyasu
1. Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;
2. Drafting the article or revising it critically for important intellectual content; and
3. Final approval of the version to be published.

STRUCTURED ABSTRACT

Objectives: Pharmacoepidemiologic safety studies are an important hypothesis-testing tool in the evaluation of drug safety issues in the postmarket period. Despite the potential to produce robust value-added data, the interpretation of findings from such studies can sometimes be challenging because of their well-recognized methodological limitations. Therefore, assessment of the quality of individual studies is essential to evaluating their credibility. The authors critically evaluated the suitability and relevance of available tools for the quality assessment of these studies.

Design: To examine the utility of individual quality assessment tools for the evaluation of observational pharmacoepidemiologic safety studies, we created an a priori assessment framework consisting of domains that include reporting elements (RE) and quality assessment attributes (QAA). Our comprehensive literature search identified distinct assessment tools and the percentage of tools assessing the pre-specified elements and attributes was tabulated.

Results: A total of 61 tools were selected for review. . Most of the tools reviewed were not designed to evaluate pharmacoepidemiologic safety studies. More than 50% considered reporting elements under the research aims, analytical approach, outcome definition and ascertainment, study population, and exposure definition and ascertainment domains. Reporting elements under the discussion and interpretation, results and study team domains were considered in less than 40% and except for the data source domain, quality attributes were considered in less than 50% of the tools.

Conclusions: Critical issues and research gaps identified include: (1) many tools do not include critical assessment elements relevant to observational pharmacoepidemiologic safety studies; (2) most do not distinguish between reporting elements and quality assessment attributes; and (3) a lack of considerations on the relative weights of different domains and elements. The development of a quality assessment tool would facilitate consistent, objective, and evidence-based assessments of pharmacoepidemiologic safety studies.

Article Summary

Article focus

This article reviews the suitability and relevance of available tools for the assessment of the quality of pharmacoepidemiologic safety studies

Key messages

1. In the context of regulatory safety-related decision making, quality assessment (i.e., assessment of the risk of bias), informs the evaluation of available evidence and enhances the appropriate utilization of available evidence in assessing the balance between benefits and risks of drugs.
2. The development of a consolidated reporting and quality assessment tool would enhance the consistent, transparent and objective evaluation of pharmacoepidemiologic safety studies. If a tool is developed, it is important to determine if there is a need for tools tailored for specific study designs or if one tool that consolidates these considerations might be helpful.
3. Key findings from our review of quality assessment tools include:

Many available quality assessment tools do not include critical assessment elements that are specifically relevant to pharmacoepidemiologic safety studies;

Most tools do not distinguish between reporting elements and quality assessment attributes; and oThere is a lack of reported considerations on the relative weights to assign to different domains and elements with respect to assessing the quality of these studies.

Strengths of the review
A priori creation of a pharmacoepidemiologic safety study assessment framework
Comprehensive review of the literature
Importance for safety-related regulatory decision making
Potential to leverage other comparable efforts in the comparative effectiveness research arena

Limitations of the review
The purpose and scope of the reviewed tools varied greatly
Each tool was reviewed by one reviewer
Some very new guidelines or checklists may have been published after our review

INTRODUCTION

Several sources of evidence on drug safety issues inform FDA postmarketing safety-related regulatory decisions, including spontaneous case reports, registries, observational pharmacoepidemiologic studies, randomized controlled trials (RCTs), meta-analyses, and other sources. Despite the well known strengths of RCTs in the assessment of drug efficacy, specific issues related to the design, methodology and transparency of experimental studies may limit their ability to fully characterize the safety profile of drugs after marketing approval.[1-4]

Pharmacoepidemiologic studies, typically observational in nature, represent an important hypothesis-testing mechanism in the evaluation of drug safety issues suspected at the time of approval and for new signals emerging in the postmarket period. In contrast to RCTs, such studies, which typically employ broader inclusion and fewer exclusion criteria and leverage claims or electronic medical record data, might better reflect the real life experience of patients. Furthermore, pharmacoepidemiologic studies afford the ability to investigate rare drug-related adverse effects, examine risks in patient subpopulations, and assess long-term adverse events.

Recent health-related legislation will increase the availability and adoption of electronic healthcare data for such studies.[5,6]

Despite the potential of pharmacoepidemiologic safety studies to produce robust value-added data, the interpretation of findings from such studies can sometimes be challenging because of their well-recognized methodological limitations, including various sources of bias and confounding.[7] These limitations also apply to the increasing number of comparative effectiveness epidemiologic studies.[5;8-10] The Institute of Medicine's recently published report highlights the importance of evaluating the quality of evidence and the significant scientific disagreements that have ensued over the quality of studies.[11] Therefore, assessment of the quality of individual studies is essential to evaluating their credibility. Transparency in reporting on the design, conduct, analysis and results of these studies is a prerequisite for the assessment of the quality of the evidence; it is first necessary to understand the relevant aspects of the study design, conduct, and analysis, along with the underlying assumptions and rationale behind the key scientific decisions undertaken by the study team to adequately evaluate the credibility of the study.[12]

The FDA recently published a draft guidance on the design, conduct, and reporting of pharmacoepidemiologic safety studies using electronic healthcare data that is designed both to enhance the transparency of reporting of such studies and to encourage critical scientific thinking regarding their design and conduct.[13] In the future, this guidance may improve the credibility of submitted pharmacoepidemiologic studies by shedding light on the pertinent aspects of studies needed to inform the evaluation of the internal and external validity of their findings. However, even if the bar for transparency and reporting of these studies is raised, it will still be necessary to evaluate the contribution of these studies to the available evidence on an emerging drug safety

issue. The Grading of Recommendations Assessment, Development and Evaluation (GRACE) evidence assessment framework, used by clinical guideline developers, appropriately separates the initial processes of quality assessment and the weighing of evidence in the formulation of guideline recommendations.[14,15] In the context of regulatory decision making concerning safety issues, the use of quality assessment tools to assess of the risk of bias may add a measure of objectivity to the scientific judgment of the available evidence and improve the quality of decision making. The benefits may not only extend to improving decision making by regulators but also by journal editors and researchers as well as potentially improving the quality of performed studies by stimulating consideration of key aspects of these studies during the development of the study approach and protocol.

Although many checklists and scales for the assessment of epidemiologic studies exist,[16-18] most are not specifically designed to evaluate pharmacoepidemiologic safety studies. **Importantly, although the principles of epidemiology apply across different fields, there are unique challenges in the design, conduct and evaluation of epidemiologic studies of unintended drug harms that warrant consideration of developing a specific validated assessment tool (e.g., confounding by indication is an important challenge that is unique to epidemiologic studies of drug effects).** Recent articles have suggested the need to develop tools for assessing the quality of these studies.[19-22] A recent publication found that systematic reviewers and meta-analysts are misusing reporting tools like STROBE due to the dearth of validated assessment instrumentss.[23]

The main objective of this article is to critically evaluate the suitability and relevance of available tools for the assessment of pharmacoepidemiologic safety studies. The ultimate goal is to stimulate discussion in the scientific community about the need for specific tools to facilitate

the transparent, objective and consistent evaluation of study quality to inform safety-related regulatory decision making.

METHODS

For purposes of this paper, quality assessment tools are defined as qualitative checklists and/or quantitative scales designed to facilitate assessment of the quality of epidemiologic studies.

A priori quality assessment framework

To examine the utility of individual quality assessment instruments for the evaluation of pharmacoepidemiologic safety studies, we created an a priori assessment framework, consisting of domains that include reporting elements (RE), and quality assessment attributes (QAA) (Table 1). Based on the expert opinion of senior FDA epidemiologists, concepts drawn from the FDA draft guidance on such studies, and key findings from seminal reviews and tools,[13;24-28] we established the domains pertaining to the design, conduct, and analysis of pharmacoepidemiologic safety studies. Within each domain we listed critical elements that need to be considered for assessing the validity and interpretation of findings from such studies. We made a distinction between the reporting elements (RE) and quality assessment attributes (QAA) for each domain. This is an important distinction as some guidelines are strictly developed to discern and evaluate reporting whereas other tools are developed to evaluate quality, which requires assessment of reporting. The selected elements and attributes presented in this Table are not intended to represent an all-inclusive list of factors, but rather to represent critical aspects impacting the internal and external validity of pharmacoepidemiologic safety studies. Of note, although the quality assessment attributes necessarily involve some subjectivity, their inclusion in an assessment tool would facilitate the consistent and objective consideration and evaluation

of key quality attributes across individual studies. As the GRADE framework developers have emphasized, although quality assessment is fundamentally subjective,[29] developing a transparent, consistent approach to assessment of quality is important, especially in the regulatory and clinical arena as patients, healthcare professionals, and sponsors benefit from consistent and transparent assessment of available evidence for use in decision making.

Literature search

A comprehensive literature search of quality assessment checklists and scales was performed in MEDLINE, EMBASE, and Web of Science. Search terms included: "assessment," "tools," "quality," "medical research," "evidence based research," "evidence based medicine," "meta-analysis," "randomized controlled trials," "biological product," "drug," "pharmaceutical preparation," "biological therapy," "bias," and "epidemiology." A total of 54 references were retrieved from this search. Two independent reviewers identified potentially relevant abstracts (n=26) from the initial literature review (inter-rater reliability > 0.85). Inclusion criteria included quality assessment tools or relevant reviews developed to evaluate RCTs, observational studies, or meta-analyses. Exclusion criteria consisted of clinical assessment tools, general articles or guidance on quality assessment, and instrumentsor reviews focused strictly on reporting and not addressing quality assessment. After reviewing each article, 10 relevant tools and 3 seminal reviews were identified; Thirteen were excluded based on the exclusion criteria above.

The most recent, relevant review articles and some individual tools for assessing the quality of epidemiologic studies were identified.[25-27] The 2007 Sanderson review,[25] the most recent, comprehensive review of instruments for assessing quality of epidemiologic studies, served as the starting point..

We also performed Google Scholar searches to identify relevant tools that might not be captured in the aforementioned search strategies. Google searches based on the first 50 hits included the following terms: "tool quality bias"; "quality assessment"; "pharmacoepidemiology"; "quality assessment epidemiology"; "tool quality assessment study"; and "scale quality assessment observational studies." Furthermore, we identified and included an EMA methodological checklist[24] because, although it is a reportingchecklist, it includes domains and considerations designed to inform safety evaluations made at a drug regulatory agency.

The assessment tools identified were reviewed by investigators based on a priori assessment criteria shown in Table 1. The percentage of tools assessing the pre-specified elements and attributes within domains was tabulated. During our review, we documented which tools employed some method of validation.

RESULTS

Overall, out of 104 tools identified, a total of 96 distinct assessment tools, including 82 from the Sanderson review, 6 from the initial literature review, 7 from the Google search, and 1 regulatory checklist (ENCEPP checklist) were considered for review (Figure 1). Out of these, 61 were selected for the in-depth review.[14;24;30-32;32;32-75;76-87] Tools exclusively focused on randomized controlled trials, tools focused on clinical assessments, and tools that did not include an explicit assessment framework were excluded (n=35).

Representation of a priori assessment domains and elements within tools

The proportion of reviewed tools that included reporting elements (RE) and quality assessment attributes (QAA) according to each a-priori defined domain within the framework is shown in Figure 1. Table 1 depicts the detailed results of our review of the domains, elements

and attributes. We highlighted the representation of select RE and QAA under each domain that may have important implications for the assessment of a pharmacoepidemiologic safety study. RE and QAA related to research aims were addressed in 69% (42/61) and 34% (21/61) of the tools, respectively. Regarding the domain assessing study population and data sources, 84% (51/61) of the tools included RE and 57% (35/61) included QAA (Table 1).

61% (37/61) of the tools included RE and 31% (19/61) included QAA under the exposure definition and ascertainment domain (Table 1). With respect to outcome definition and ascertainment domain, 69% (42/61) of the tools included RE and 36% (22/61) included QAA (Table 1). Out of the 61 reviewed tools, 85% (52/61) and 49% (30/61) included RE and QAA under the analytic approach domain (Table 1). Under the results domain, only 36% (22/61) and 7% (4/61) of tools included RE and QAA respectively (Table 1).

Of the 61 reviewed tools, 36% (22/61) and 20% (12/61) of tools included RE and QAA under the discussion and interpretation domain (Table 1). 7% (4/61) of the tools addressed the description of the study team (RE) and the independence of team and funding sources (QAA).

More than half of the reviewed instruments considered reporting elements for domains including research aims, analytical approach, outcome definition and ascertainment, study population, and exposure definition and ascertainment. Domains related to the discussion and interpretation, results and study team were considered in less than 40% of the reviewed tools. With the exception of the study population/data sources domain, QAA research domains were considered in less than half of the assessment tools, with less than 10% considering results and study team domains. Many did not address all pertinent domains.

Most reviewed checklists and scales were not specifically designed to assess epidemiologic studies of drug-related harms. Although the EMA framework was designed to

increase transparency of pharmacoepidemiologic studies, it focuses on reporting versus quality assessment. Our review constitutes the only recent comprehensive review of available assessment tools to determine if any are appropriate and sufficient for the evaluation of pharmacoepidemiologic studies. Only a small number of the reviewed instruments employed some method of validation.[30-36] Most of the tools did not differentiate between reporting elements (RE) and quality assessment attributes (QAA) whereas others stratified by these aspects. A small number included distinct assessment criteria for different epidemiologic study designs (e.g. case-control, cohort). Tools focused more on RE than QAA. Figure 1 displays the percentage of checklists and scales that included criteria on the assessment domains and elements.

The proportion of reviewed tools that included reporting elements (RE) and quality assessment attributes (QAA) according to each a-priori defined domain within the framework is shown in Figure 1.

DISCUSSION

Based on our review, there is no specific tool that is adequately designed for the robust evaluation of pharmacoepidemiologic studies of drug safety. No single tool considered all the selected domains and elements and most tools failed to address critical evaluation elements. Making a distinction between RE and QAA is important as even if an element of the study is mentioned in the final report, one must then determine if this was appropriate for the specific study in the context of the drug safety question. Only a fewinstruments specifically made this distinction as we did in our a priori framework. Additionally, important RE and QAA were lacking in most of the checklists and scales we reviewed which highlights the need for a tool

focused on the evaluation of epidemiologic studies designed to evaluate drug-related harms; this need has been previously identified by others.[88]

Quality attributes related to exposure definition and ascertainment were considered in less than half of the assessment tools, with less than 15% including RE and QAA pertaining to the comparator group, despite the fact that the selection of a comparator is critical for drug safety and effectiveness trials and epidemiologic studies, as the choice of suboptimal comparators can provide misleading results.[89;90] Only 30% of the instrument included quality assessment elements pertaining to the validity and appropriateness of the operational aspects of exposure ascertainment, and only 36% addressed quality attributes of validation of outcome ascertainment approaches. These are important facets of pharmacoepidemiologic safety studies their misclassification may lead to false negative findings regarding the association between a drug and adverse event.

Only about 40% of the checklists and scales included QAA on approaches to handle confounding and biases. As observational studies are not randomized, the approaches to handle confounding and bias are of paramount importance.[7;91;92] This is an important limitation of most tools because there are often uncertainties regarding results from pharmacoepidemiologic studies due to the limitations of electronic healthcare data and complex nature of the practice of medicine.[92] Only a small percentage of tools (28% RE; 18% QAA) included elements on the consideration of study findings in the context of the design, conduct, limitations, and statistical power despite the fact that these elements are essential in assessing implications of study findings.

Some of the tools we reviewed were designed as "all-purpose" assessment instruments for evaluation of clinical trials and observational studies, while others focused on a particular

study design (e.g., case-control, cohort). It may be useful to create one consolidated, validated tool for evaluating observational pharmacoepidemiologic safety studies focused on general reporting and quality attributes; tools for the specific study designs, i.e., case-control and cohort studies, may also be useful due to some of the unique aspects of these designs. By creating such a tool, regulatory agencies, clinical guideline developers, and clinicians could consistently evaluate studies and for decision making. The creation of this instrument could be led by an independent group of expert methodologists, perhaps with input from multiple stakeholders, including regulatorsand professional organizations.

Although we did not address weighing of importance of different domains and elements based on their relative impact on study contribution to the available streams of evidence, this may be an important consideration in the formulation of an assessment tool. Also, it is not clear if numerical scores are helpful in assessing the quality of epidemiologic studies, as when numerical scores were used to evaluate systematic reviews or meta-analyses of such studies, they did not produce valid results.[93] The appropriate tradeoff between the utility of a checklist or scalefor review and the comprehensiveness of the evaluation elements has yet to be determined. This is complicated by the lack of validation of most of the available tools. Before these issues can be addressed, it is first necessary to engage in a broader discussion of the utility of such assessment tools in the evaluation of pharmacoepidemiologic safety studies. It is worth noting that critical assessment elements of pharmacoepidemiologic studies focused on effectiveness may be different than those focused on safety; however, pharmacoepidemiologic comparative effectiveness studies focus on both comparative safety and benefits associated with drugs and thus such elements are not mutually exclusive.[94] Thus, it is important to consider the potential to

leverage current efforts to create a validated assessment tool (GRACE checklist[95]) for observational comparative effectiveness pharmacoepidemiologic studies.

Our review has some limitations. The purpose and scope of the checklists and scaleswe reviewed varied greatly. Although we conducted a comprehensive review, there may be tools that we were unable to access or that were published after our search. If a reporting or quality assessment element contained some aspects of the element, we counted this as full representation, even if not all the important sub-elements were included. Each checklist or scale was reviewed by one study team member; repeating the evaluations via a second reviewer was deemed unnecessary at this stage as the primary goal of the review was to obtain a broad understanding of the utility of available assessment tools in evaluating pharmacoepidemiologic safety studies based on a preliminary assessment framework. Some factors that may be increasingly relevant in future studies, such as electronic health records with linkages to other data sources like outpatient claims, health information exchanges, or personal health records, were not included in our framework but may be included in a future validated instrument. Guidelines and checklists published after the time period of our review have included some elements that may be important for future linked studies which may leverage the increasing availability of these data sources.[96]

In the evaluation of many emerging safety issues, pharmacoepidemiologic safety studies are discussed and may influence safety-related decision making. However, often the quality-driven contribution of such studies is not discussed in a consistent way. The development of an assessment tool based on expert input may facilitate consistent, evidence-based quality assessment of such studies and the subsequent determination of their value based on evaluating the impact of bias on the robustness of a study results, and the interpretation of its findings,

within the context of the specific drug safety issue. The framework we developed may serve as a foundation for future development of such an instrument. Efforts to improve the evaluation of the contribution of pharmacoepidemiologic safety studies would be consistent with the FDA's focus on strengthening regulatory safety science.[97] If after further consideration and discussions with stakeholders development of a tool to evaluate epidemiologic data for drug safety is pursued, it would be necessary to first determine the scope of the assessment tool as well as steps for its comprehensive validation. Further, relevant aspects of the design and analysis of pharmacoepidemiology studies should be considered (we refer the reader to some helpful references, 98, 99, 100).[98-100] Importantly, such a tool would be intended to complement, and not replace, expert clinical, methodological, and statistical expertise necessary to complete a robust evaluation and determination of the contribution of a specific pharmacoepidemiologic safety study to the available evidence for regulatory decision making.

Funding Statement

This research received no specific funding.

Table 1: Reporting elements and quality assessment attributes according to selected domains and percent representation among reviewed tools

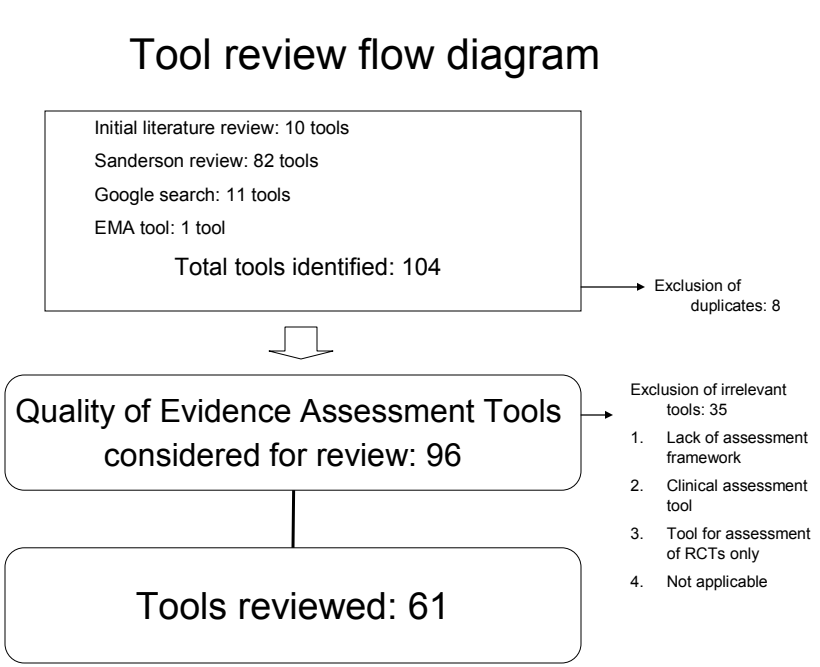
Reporting Elements (REs):	Percent representation	Quality Assessment Attributes (QAAs):	Percent representation
A. Research aims	69%		34%
RE 1: Description of study objectives, research aims, design, study population and data source, exposure, and outcome	69%	QAA 1: Appropriateness of pre-specified aims, design, population, exposure, and outcome to address research aim	34%
B. Study population: data sources	84%		57%
RE 1: Description of participation rates and discontinuation rates	77%	QAA 1: Extent of participation rates and discontinuation rates	56%
RE 2: Description of denominator used for risk assessment	11%	QAA 2: Appropriateness of denominator used for risk assessment	3%
C. Exposure definition and ascertainment	61%		31%
RE 1: Description of operational aspects of exposure ascertainment and definition	49%	QAA 1: Validity and appropriateness of operational aspects used to ascertain and define exposure status	30%
RE 2: Description of blinding of outcome status	21%		
RE 3: Selection of exposure risk window	5%	QAA 2: Appropriateness of selected exposure risk window	3%
RE 4: Description of selected type of users (incident v. prevalent users)	0%	QAA 3: Appropriateness of selected numerator for risk assessment	0%
RE 5: Description of comparison group	10%	QAA 4: Appropriateness of comparison group	5%
D. Safety outcome definition and ascertainment	69%		36%
RE 1: Description of operational aspects of outcome ascertainment and definition	51%	QAA 1: Appropriateness/validity of outcome ascertainment strategies and outcome definition	33%
RE 2: Description of blinding of exposure status from those ascertaining/validating outcomes	25%		
RE 3: Description of follow up time	16%	QAA 2: Adequacy of follow up time to address research question	13%
RE 4: Description of composite outcome, if relevant	0%	QAA 3: Adequacy of composite safety outcome, if relevant	0%
E. Analytical approach	85%		49%
RE 1: Description of analytic approach, including approaches to handle confounding and biases	80%	QAA 1: Appropriateness of described analytic approach	26%
RE 2: Description of a priori sample size/power calculations	44%	QAA 2: Appropriateness of approaches to handle confounding and biases	39%

RE 3: Description of data integration methods, when relevant	3%	QAA 3: Description of a priori sample size/power calculations	21%
RE 4: Description of measures of frequency and association	7%	QAA 4: Appropriateness of data integration methods, when relevant	0%
RE 5: Description of a priori specifications of subgroup analyses	5%		
F. Results	36%		7%
RE 1: Description of main results (unadjusted and adjusted estimates and confidence intervals) and sensitivity analyses	25%	QAA 1: Consistency of primary, secondary, and sensitivity analyses and consistency of confounding effects with known associations	2%
RE 2: Description of patient disposition	15%	QAA 2: Impact of patient disposition on study integrity and generalizability of findings	6%
RE 3: Description of characteristics of population by comparison group	18%		
G. Discussion and interpretation	36%		20%
RE 1: Description of findings in relation to pertinent issues related to study design, conduct, limitations, and power	28%	QAA 1: Consideration of findings in relation to pertinent issues related to study design, conduct, limitations, and power	18%
RE 2: Description of plausibility of findings and clinical significance and discussion/exploration of alternative explanations, comparison with other findings	21%	QAA 2: Discussion of plausibility of findings and clinical significance and discussion of alternative explanations, comparison with other findings	11%
H. Study team	7%		3%
RE 1: Description of study team, conflict of interest, funding sources	7%	QAA 1: Relevance of study team credentials and experience to the research area	0
		QAA 2: Independence of study team and funding sources	3%

Figure 1 Percent representation by domain of reporting and quality aspects considered by the assessment tools

INSERT FIGURE ONE HERE

Appendix: Flow diagram



Reference List

(1) Hammad TA, Pinheiro SP, Neyarapally GA. Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clin Trials* 2011;8:559-570.

(2) Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737-1739.

(3) Pitrou I, Boutron I, Ahmad N, et al. Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756-1761.

(4) Clark DW, Coulter DM, Besag FM. Randomized controlled trials and assessment of drug safety. *Drug Saf* 2008;31:1057-1061.

(5) American Recovery and Reinvestment Act. Pub.L. 111-5. 2009.

(6) Bruen BK, Ku L, Burke MF, et al. More than four in five office-based physicians could qualify for federal electronic health record incentives. *Health Aff (Millwood)* 2011;30:472-480.

(7) Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323-337.

(8) Dreyer N, Tunis S, Berger M. Why Observational Studies Should Be Among The Tools Used In Comparative Effectiveness Research. *Health Affairs* 2010;29:1818.

(9) Dreyer N. Do We Need Good Practice Principles for Observational Comparative Effectiveness Research? *GRACE Initiative* [serial online] 2009.

(10) Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. *Stat Med* 2010;29:1963-1976.

(11) Institute of Medicine. Ethical and Scientific Issues in Studying the Safety of Approved Drugs. 2012.

(12) von EE, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344-349.

(13) FDA. FDA Draft Guidance: Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data Sets. FDA Draft Guidance . 2011.

(14) Atkins D, Best D, Briss PA et al. Grading quality of evidence and strength of recommendations. *BMJ* 2004;%19;328:1490.

(15) Guyatt GH, Oxman AD, Vist GE et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924-926.

(16) Deeks JJ, Dinnes J, D'Amico R et al. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-173.

(17) Sanderson S, Tatt ID, Higgins JP. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666-676.

(18) Lohr KN. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care* 2004;16:9-18.

(19) Cornelius VR, Perrio MJ, Shakir SA, et al. Systematic reviews of adverse effects of drug interventions: a survey of their conduct and reporting quality. *Pharmacoepidemiol Drug Saf* 2009;18:1223-1231.

(20) Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 2005;142:1090-1099.

(21) Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.:32.

(22) Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol* 2006;59:765-769.

(23) da Costa BR, Cevallos M, Altman DG, et al. Uses and misuses of STROBE statement: bibliographic study. *BMJ Open* 2011;doi:10.1136/bmjopen-2010-000048 .

(24) EMA. Checklist for Methodological Standards for ENCePP Study Protocols. 2010.

(25) Sanderson. Tools for assessing the quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International Journal of Epidemiology* 2007;36:666.

(26) Deeks. Evaluating non-randomised intervention studies. *Health Technology Assessment* 2003;7.

(27) Lohr K. Rating the strength of scientific evidence: relevance for quality improvement programs. *International Journal for Quality in Health Care* 2004;16:9-18.

(28) von Elm E. STROBE. *Journal of Clinical Epidemiology* 2008;344.

(29) Balshem H, Helfand M, Schunemann HJ et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-406.

(30) Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health* 1998;52:377-384.

(31) Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of oro-facial pain. *J Dent* 2001;29:451-467.

(32) Margetts BM, Thompson RL, Key T et al. Development of a scoring system to judge the scientific quality of information from case-control and cohort studies of nutrition and disease. *Nutr Cancer* 1995;24:231-239.

(33) Rangel SJ, Kelsey J, Colby CE, et al. Development of a quality assessment scale for retrospective clinical studies in pediatric surgery. *J Pediatr Surg* 2003;38:390-396.

(34) Genaidy AM, Lemasters GK, Lockey J et al. An epidemiological appraisal instrument - a tool for evaluation of epidemiological studies. *Ergonomics* 2007;50:920-960.

(35) Santaguida, Raina. McMaster Quality Assessment Scale of Harms (McHarm) for primary studies . 2008.

(36) Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994;272:101-104.

(37) Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. *J Clin Nurs* 1994;3:271-277.

(38) Cameron I, Crotty M, Currie C et al. Geriatric rehabilitation following fractures in older people: a systematic review. *Health Technol Assess* 2000;4:i-111.

(39) Carneiro AV. Critical appraisal of prognostic evidence: practical rules. *Rev Port Cardiol* 2002;21:891-900.

(40) NHS. Critical Appraisal Skills Programme: appraisal tools. Public Health Resource Unit N, editor. 2003.

(41) Center for Occupational and Environmental Health. Critical appraisal. School of Epidemiology and Health Sciences UoM, editor. 2003.

(42) DuRant RH. Checklist for the evaluation of research articles. *J Adolesc Health* 1994;15:4-8.

(43) Elwood M. Forward projection--using critical appraisal in the design of studies. *Int J Epidemiol* 2002;31:1071-1073.

(44) Esdaile JM, Horwitz RI. Observational studies of cause-effect relationships: an analysis of methodologic problems as illustrated by the conflicting data for the role of oral contraceptives in the etiology of rheumatoid arthritis. *J Chronic Dis* 1986;39:841-852.

(45) Gardner MJ, Machin D, Campbell MJ. Use of check lists in assessing the statistical content of medical studies. *Br Med J (Clin Res Ed)* 1986;292:810-812.

(46) Health Evidence Bulletin. Questions to assist with the critical appraisal of an observational study eg cohort, case-control, cross-sectional. Wales, editor. 2004.

(47) Levine M, Walter S, Lee H, et al. Users' guides to the medical literature. IV. How to use an article about harm. Evidence-Based Medicine Working Group. *JAMA* 1994;271:1615-1619.

(48) Lichtenstein MJ, Mulrow CD, Elwood PC. Guidelines for reading case-control studies. *J Chronic Dis* 1987;40:893-903.

(49) Federal Focus I. The London Principles for Evaluating Epidemiologic Data in Regulatory Risk Assessment. *Federal Focus* [serial online] 2004.

(50) Margetts BM, Vorster, Venter. Evidence-Based Nutrition - Review of Nutritional Epidemiological Studies. *South African Journal of Clinical Nutrition* 2002;15:68-73.

(51) University of Montreal. Critical Appraisal Worksheet. 2004.

(52) Mulrow CD, Lichtenstein MJ. Blood glucose and diabetic retinopathy: a critical appraisal of new evidence. *J Gen Intern Med* 1986;1:73-77.

(53) Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.:25.

(54) Scottish Intercollegiate Guidelines Network. SIGN 50: A guideline developer's handbook. *Scottish Intercollegiate Guidelines Network* [serial online] 2004.

(55) Parker S. Critical appraisal: Guidelines for the critical appraisal of a paper. *Website* [serial online] 2004.

(56) Zaza S, Wright-De Aguero LK, Briss PA et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000;18:44-74.

(57) Zola P, Volpe T, Castelli G et al. Is the published literature a reliable guide for deciding between alternative treatments for patients with early cervical cancer? *Int J Radiat Oncol Biol Phys* 1989;16:785-797.

(58) Ariens GA, van MW, Bongers PM, et al. Physical risk factors for neck pain. *Scand J Work Environ Health* 2000;26:7-19.

(59) Bhutta AT, Cleves MA, Casey PH, et al. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 2002;288:728-737.

(60) Borghouts JA, Koes BW, Bouter LM. The clinical course and prognostic factors of non-specific neck pain: a systematic review. *Pain* 1998;77:1-13.

(61) Carson CA, Fine MJ, Smith MA, et al. Quality of published reports of the prognosis of community-acquired pneumonia. *J Gen Intern Med* 1994;9:13-19.

(62) Loney PL, Chambers LW, Bennett KJ, et al. Critical appraisal of the health research literature: prevalence or incidence of a health problem. *Chronic Dis Can* 1998;19:170-176.

(63) Corrao G, Bagnardi V, Zambon A, et al. Exploring the dose-response relationship between alcohol consumption and the risk of several alcohol-related conditions: a meta-analysis. *Addiction* 1999;94:1551-1573.

(64) Garber BG, Hebert PC, Yelle JD, et al. Adult respiratory distress syndrome: a systemic overview of incidence and risk factors. *Crit Care Med* 1996;24:687-695.

(65) Goodman SN, Berlin J, Fletcher SW, et al. Manuscript quality before and after peer review and editing at Annals of Internal Medicine. *Ann Intern Med* 1994;121:11-21.

(66) Jabbour M, Osmond MH, Klassen TP. Life support courses: are they effective? *Ann Emerg Med* 1996;28:690-698.

(67) Kreulen CM, Creugers NH, Meijering AC. Meta-analysis of anterior veneer restorations in clinical studies. *J Dent* 1998;26:345-353.

(68) Krogh CL. A checklist system for critical review of medical literature. *Med Educ* 1985;19:392-395.

(69) Littenberg B, Weinstein LP, McCarren M et al. Closed fractures of the tibial shaft. A meta-analysis of three methods of treatment. *J Bone Joint Surg Am* 1998;80:174-183.

(70) Longnecker MP, Berlin JA, Orza MJ, et al. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988;260:652-656.

(71) Manchikanti L, Singh V, Vilims BD, et al. Medial branch neurotomy in management of chronic spinal pain: systematic review of the evidence. *Pain Physician* 2002;5:405-418.

(72) Meijer R, Ihnenfeldt DS, van LJ, et al. Prognostic factors in the subacute phase after stroke for the future residence after six months to one year. A systematic review of the literature. *Clin Rehabil* 2003;17:512-520.

(73) Nguyen QV, Bezemer PD, Habets L, et al. A systematic review of the relationship between overjet size and traumatic dental injuries. *Eur J Orthod* 1999;21:503-515.

(74) Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989;84:815-827.

(75) Stock SR. Workplace ergonomic factors and the development of musculoskeletal disorders of the neck and upper limbs: a meta-analysis. *Am J Ind Med* 1991;19:87-107.

(76) van der Windt DA, Thomas E, Pope DP et al. Occupational risk factors for shoulder pain: a systematic review. *Occup Environ Med* 2000;57:433-442.

(77) Bollini P, Garcia Rodriguez LA, Perez GS, et al. The impact of research quality and study design on epidemiologic estimates of the effect of nonsteroidal anti-inflammatory drugs on upper gastrointestinal tract disease. *Arch Intern Med* 1992;152:1289-1295.

(78) Ciliska D, Hayward S, Thomas H et al. A systematic overview of the effectiveness of home visiting as a delivery strategy for public health nursing interventions. *Can J Public Health* 1996;87:193-198.

(79) Cowley DE. Prostheses for primary total hip replacement. A critical appraisal of the literature. *Int J Technol Assess Health Care* 1995;11:770-778.

(80) Effective Public Health Practice Project. Quality Assessment Tools for Quantitative Studies. 2003.

(81) Faculty of Medical and Health Sciences. EPIQ (Effective Practice, Informatics, and Quality Improvement). 2004. University of Auckland.

(82) Fowkes FG, Fulton PM. Critical appraisal of published research: introductory guidelines. *BMJ* 1991;302:1136-1140.

(83) Gyorkos TW, Tannenbaum TN, Abrahamowicz M et al. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994;85 Suppl 1:S8-13.:S8-13.

(84) Spitzer WO, Lawrence V, Dales R et al. Links between passive smoking and disease: a best-evidence synthesis. A report of the Working Group on Passive Smoking. *Clin Invest Med* 1990;13:17-42.

(85) Steinberg EP, Eknoyan G, Levin NW et al. Methods used to evaluate the quality of evidence underlying the National Kidney Foundation-Dialysis Outcomes Quality Initiative Clinical Practice Guidelines: description, findings, and implications. *Am J Kidney Dis* 2000;36:1-11.

(86) El BN, Middel B, van Dijk JP, et al. Are the outcomes of clinical pathways evidence-based? A critical appraisal of clinical pathway evaluation research. *J Eval Clin Pract* 2007;13:920-929.

(87) Salford University. Evaluation Tool for Quantitative Research Studies. 2001.

(88) Chou R, Aronson N, Atkins D et al. AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. *J Clin Epidemiol* 2010;63:502-512.

(89) Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff (Millwood)* 2010;29:1842-1848.

(90) Psaty BM, Weiss NS. NSAID trials and the choice of comparators--questions of public health importance. *N Engl J Med* 2007;356:328-330.

(91) Strom B. *Pharmacoepidemiology*. 4th ed. John Wiley & Sons Ltd, 2005.

(92) Brookhart MA, Sturmer T, Glynn RJ, et al. Confounding control in healthcare database research: challenges and potential approaches. *Med Care* 2010;48:S114-S120.

(93) Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol* 2010;25:603-605.

(94) Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *JAMA* 2010;303:985-986.

(95) GRACE. GRACE Checklist. 2012.

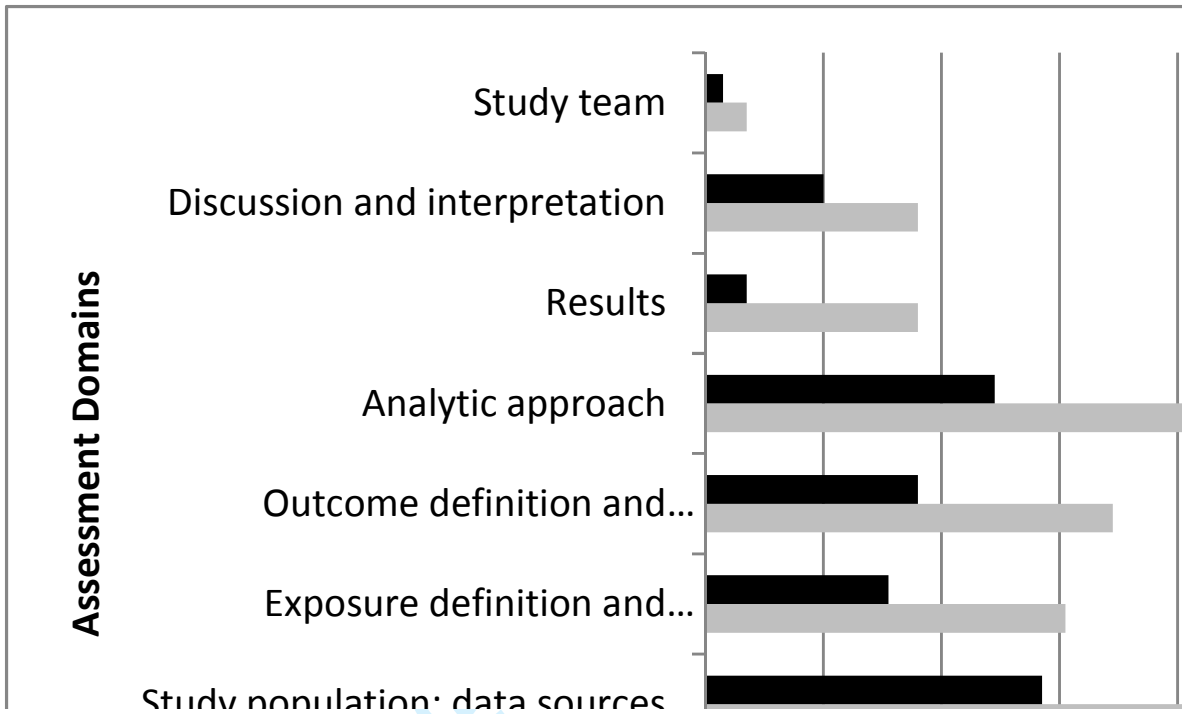
(96) Hall GC, Sauer B, Bourke A, et al. Guidelines for good database selection and use in pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf* 2012;21:1-10.

(97) Hamburg MA. Shattuck lecture. Innovation, regulation, and the FDA. *N Engl J Med* 2010;363:2228-2232.

(98) Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. *J Clin Epidemiol* 1989;42:325-331.

(99) Guess HA. Behavior of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *J Clin Epidemiol* 1989;42:1179-1184.

(100) Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858-868.



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Lit.review
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	1
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	2-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	N/A
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	N/A
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	N/A
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5-6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5 - 7
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5, table
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Table
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	N/A

PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	14 (new)
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	N/A
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	8 - 12
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	11
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	8-12
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	1

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: www.prisma-statement.org.