

# Review of quality assessment tools for the evaluation of pharmacoepidemiological safety studies

George A Neyarapally, Tarek A Hammad, Simone P Pinheiro, Solomon Iyasu

**To cite:** Neyarapally GA, Hammad TA, Pinheiro SP, *et al.* Review of quality assessment tools for the evaluation of pharmacoepidemiological safety studies. *BMJ Open* 2012;**2**:e001362. doi:10.1136/bmjopen-2012-001362

► Prepublication history and additional material for this paper are available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-001362>).

► Additional supplementary files are published online only. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2012-001362>).

Received 8 May 2012  
Accepted 21 August 2012

This final article is available for use under the terms of the Creative Commons Attribution Non-Commercial 2.0 Licence; see <http://bmjopen.bmj.com>

Office of Surveillance and Epidemiology, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA

## Correspondence to

Dr George A Neyarapally;  
[george.neyarapally@fda.hhs.gov](mailto:george.neyarapally@fda.hhs.gov)

## ABSTRACT

**Objectives:** Pharmacoepidemiological studies are an important hypothesis-testing tool in the evaluation of postmarketing drug safety. Despite the potential to produce robust value-added data, interpretation of findings can be hindered due to well-recognised methodological limitations of these studies. Therefore, assessment of their quality is essential to evaluating their credibility. The objective of this review was to evaluate the suitability and relevance of available tools for the assessment of pharmacoepidemiological safety studies.

**Design:** We created an a priori assessment framework consisting of reporting elements (REs) and quality assessment attributes (QAAs). A comprehensive literature search identified distinct assessment tools and the prespecified elements and attributes were evaluated.

**Primary and secondary outcome measures:** The primary outcome measure was the percentage representation of each domain, RE and QAA for the quality assessment tools.

**Results:** A total of 61 tools were reviewed. Most tools were not designed to evaluate pharmacoepidemiological safety studies. More than 50% of the reviewed tools considered REs under the research aims, analytical approach, outcome definition and ascertainment, study population and exposure definition and ascertainment domains. REs 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:** Many tools failed to include critical assessment elements relevant to observational pharmacoepidemiological safety studies and did not distinguish between REs and QAAs. Further, there is 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 pharmacoepidemiological safety studies.

## INTRODUCTION

Several sources of evidence on drug safety issues inform Food and Drug Administration (FDA) postmarketing safety-related regulatory

## ARTICLE SUMMARY

### Article focus

- This article reviews the suitability and relevance of available tools for the assessment of the quality of pharmacoepidemiological safety studies.

### Key messages

- In the context of regulatory safety-related decision making, quality assessment (ie, assessment of the risk of bias), informs the evaluation of available evidence and enhances the appropriate utilisation 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 pharmacoepidemiological 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 pharmacoepidemiological safety studies.
  - Most tools do not distinguish between reporting elements (REs) and quality assessment attributes (QAAs).
  - 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.

### Strengths and limitations of this study

- Strengths
  - A priori creation of a pharmacoepidemiological 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
  - The purpose and scope of the reviewed tools varied greatly.
  - Each tool was reviewed by one reviewer.

decisions, including spontaneous case reports, registries, observational pharmacoepidemiological studies, randomised 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 characterise the safety profile of drugs after marketing approval.<sup>1-4</sup> Pharmacoepidemiological 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 criteria and leverage claims or electronic medical record data, might better reflect the real-life experience of patients. Furthermore, pharmacoepidemiological 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.<sup>5,6</sup>

Despite the potential of pharmacoepidemiological safety studies to produce robust value-added data, the interpretation of findings from such studies can sometimes be challenging because of their well-recognised methodological limitations, including various sources of bias and confounding.<sup>7</sup> These limitations also apply to the increasing number of comparative effectiveness epidemiological studies.<sup>8-10</sup> 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.<sup>11</sup> 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.<sup>12</sup>

The FDA recently published a draft guidance on the design, conduct and reporting of pharmacoepidemiological 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.<sup>13</sup> In the future, this guidance may improve the credibility of submitted pharmacoepidemiological 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 (GRADE) 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.<sup>14,15</sup> In the context of regulatory decision making concerning safety issues, the use of quality assessment tools to assess 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 epidemiological studies exist,<sup>16-18</sup> most are not specifically designed to evaluate pharmacoepidemiological safety studies. Importantly, although the principles of epidemiology apply across different fields, there are unique challenges in the design, conduct and evaluation of epidemiological studies of unintended drug harms that warrant consideration of developing a specific validated assessment tool (eg, confounding by indication is an important challenge that is unique to epidemiological studies of drug effects). Recent articles have suggested the need to develop tools for assessing the quality of these studies.<sup>19-22</sup> A recent publication found that systematic reviewers and meta-analysts are misusing reporting tools like Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) due to the dearth of validated assessment instruments.<sup>23</sup>

The main objective of this study is to critically evaluate the suitability and relevance of available tools for the assessment of pharmacoepidemiological 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 epidemiological studies.

### A priori quality assessment framework

To examine the utility of individual quality assessment instruments for the evaluation of pharmacoepidemiological safety studies, we created an a priori assessment framework, consisting of domains that include reporting elements (REs), and quality assessment attributes (QAAs) (table 1). Based on the expert opinion of FDA epidemiologists, concepts drawn from the FDA draft guidance on such studies, and key findings from seminal reviews tools and documents,<sup>13,24-28</sup> we established the

**Table 1** Reporting elements (REs) and quality assessment attributes (QAAs) according to selected domains and percent representation among reviewed tools

Reporting elements (REs)	Percent representation (%)	Quality assessment attributes (QAAs)	Percent representation (%)
<b>A. Research aims</b>	69		34
RE 1: Description of study objectives, research aims, design, study population and data source, exposure and outcome	69	QAA 1: Appropriateness of prespecified aims, design, population, exposure and outcome to address research aim	34
<b>B. Study population: data sources</b>	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
<b>C. Exposure definition and ascertainment</b>	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 vs 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
<b>D. Safety outcome definition and ascertainment</b>	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
<b>E. Analytical approach</b>	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		
<b>F. Results</b>	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

Continued

Table 1 Continued

Reporting elements (REs)	Percent representation (%)	Quality assessment attributes (QAAs)	Percent representation (%)
RE 2: Description of patient disposition	15	QAA 2: Impact of patient disposition on study integrity and generalisability of findings	6
RE 3: Description of characteristics of population by comparison group	18		
<b>G. Discussion and interpretation</b>	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
<b>H. Study team</b>	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

domains pertaining to the design, conduct and analysis of pharmacoepidemiological 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 REs and QAAs 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 pharmacoepidemiological safety studies. Of note, although the QAAs 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 emphasised, although quality assessment is fundamentally subjective,<sup>29</sup> 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’, ‘randomised 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 instruments or reviews focused strictly on reporting and not addressing quality assessment. After reviewing each paper, 10 relevant tools and 3 seminal reviews were identified; 13 were excluded based on the exclusion criteria above.

The most recent, relevant review articles and some individual tools for assessing the quality of epidemiological studies were identified.<sup>25–27</sup> The 2007 Sanderson review,<sup>25</sup> the most recent, comprehensive review of instruments for assessing quality of epidemiological 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 European Medicines Agency (EMA) methodological checklist<sup>24</sup> because, although it is a reporting checklist, 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 prespecified elements and attributes within domains (percent representation) 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.<sup>14 24 30-37</sup> Tools exclusively focused on RCTs, 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 REs and QAAs 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 pharmacoepidemiological 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).

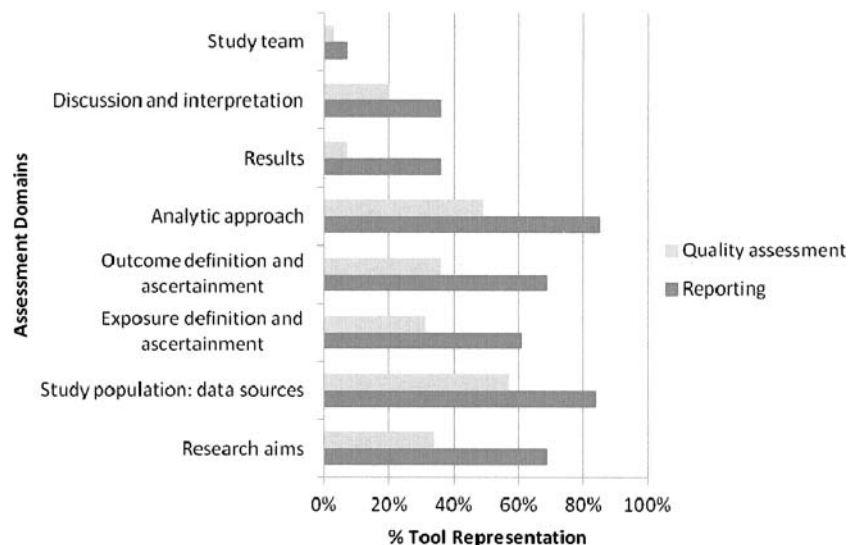
In all, 61% (37/61) of the tools included RE and 31% (19/61) included QAA under the exposure definition and ascertainment domain (table 1). With respect to outcome definition and ascertainment domain, 69% (42/61) of the tools included RE and 36% (22/61) included QAA (table 1). Out of the 61 reviewed 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). Overall, 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 REs 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 epidemiological studies of drug-related harms. Although the EMA framework was designed to increase transparency of pharmacoepidemiological 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 pharmacoepidemiological studies. Only a small number of the reviewed instruments employed some method of validation.<sup>30-36</sup> Most of the tools did

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



not differentiate between REs and QAAs whereas others stratified by these aspects. A small number included distinct assessment criteria for different epidemiological study designs (eg, 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 REs and QAAs 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 pharmacoepidemiological 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 few instruments 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 epidemiological studies designed to evaluate drug-related harms; this need has been previously identified by others.<sup>88</sup>

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 epidemiological studies, as the choice of suboptimal comparators can provide misleading results.<sup>89–90</sup> 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 pharmacoepidemiological 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 randomised, the approaches to handle confounding and bias are of paramount importance.<sup>7–91–92</sup> This is an important limitation of most tools because there are often uncertainties regarding results from pharmacoepidemiological studies due to the limitations of electronic healthcare data and complex nature of the practice of medicine.<sup>92</sup> 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 (eg, case–control, cohort). It may be useful to create one consolidated, validated tool for evaluating observational pharmacoepidemiological safety studies focused on general reporting and quality attributes; tools for the specific study designs, that is, 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 regulators and professional organisations.

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 epidemiological studies, as when numerical scores were used to evaluate systematic reviews or meta-analyses of such studies, they did not produce valid results.<sup>93</sup> The appropriate tradeoff between the utility of a checklist or scale 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 pharmacoepidemiological safety studies. It is worth noting that critical assessment elements of pharmacoepidemiological studies focused on effectiveness may be different than those focused on safety; however, pharmacoepidemiological comparative effectiveness studies focus on both comparative safety and benefits associated with drugs and thus such elements are not mutually exclusive.<sup>94</sup> Thus, it is important to consider the potential to leverage current efforts to create a validated assessment tool (GRACE checklist<sup>95</sup>) for observational comparative effectiveness pharmacoepidemiological studies.

Our review has some limitations. The purpose and scope of the checklists and scales 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 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 pharmacoepidemiological

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.<sup>96</sup>

In the evaluation of many emerging safety issues, pharmacoepidemiological 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 pharmacoepidemiological safety studies would be consistent with the FDA's focus on strengthening regulatory safety science.<sup>97</sup> If after further consideration and discussions with stakeholders development of a tool to evaluate epidemiological 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).<sup>98–100</sup> 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 pharmacoepidemiological safety study to the available evidence for regulatory decision making.

**Acknowledgements** The authors would like to thank Gerald Dal Pan and Judy Staffa for their critical evaluation of the paper and thoughtful suggestions.

**Contributors** AGAN was involved in Substantial contributions to the conception and design, acquisition of data or analysis and interpretation of the data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

TAH was involved in Substantial contributions to the conception and design, acquisition of data or analysis and interpretation of the data; drafting the article or revising it critically for important intellectual content; and Final approval of the version to be published.

SPP was involved in Substantial contributions to the conception and design, acquisition of data or analysis and interpretation of the data;

drafting the article or revising it critically for important intellectual content; and final approval of the version to be published. DSI was involved in Substantial contributions to the conception and design, acquisition of data, or analysis and interpretation of the data;

drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

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

**Competing interests** None.

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

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** No additional data are available.

## REFERENCES

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;8:559–70.
2. Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737–9.
3. Pitrou I, Boutron I, Ahmad N, *et al.* Reporting of safety results in published reports of randomized controlled trials. *Arch Intern Med* 2009;169:1756–61.
4. Clark DW, Coulter DM, Besag FM. Randomized controlled trials and assessment of drug safety. *Drug Saf* 2008;31:1057–61.
5. American Recovery and Reinvestment Act. Pub.L. 2009;111–5
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;30:472–80.
7. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
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. [http://www.graceprinciples.org/art/GRACE\\_principles\\_25\\_october\\_2009.pdf](http://www.graceprinciples.org/art/GRACE_principles_25_october_2009.pdf) (accessed on 3 Mar 2011).
10. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: policy context, methods development and research infrastructure. *Stat Med* 2010;29:1963–76.
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;61:344–9.
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;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;336:924–6.
16. Deeks JJ, Dinnes J, D'Amico R, *et al.* Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:173.
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;36:666–76.
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;18:1223–31.
20. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med* 2005;142:1090–9.
21. Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. *BMC Med Res Methodol* 2007;7:32.
22. Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. *J Clin Epidemiol* 2006;59:765–9.
23. da Costa BR, Cevallos M, Altman DG, *et al.* Uses and misuses of STROBE statement: bibliographic study. *BMJ Open* 2011; 1:e000048 doi:10.1136/bmjopen-2010-000048.
24. EMA. Checklist for Methodological Standards for ENCePP Study Protocols. 2010.

25. Sanderson S, Tatt ID, Higgins JP. Tools for assessing the quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *Int J Epidemiol* 2007;36:666.
26. Deeks JJ, Dinnes J, D'Amico R. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:iii-x, 1-173.
27. Lohr K. Rating the strength of scientific evidence: relevance for quality improvement programs. *Int J Qual Health Care* 2004;16:9-18.
28. von Elm E. STROBE. *J Clin Epidemiol* 2008;61:344-9.
29. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401-6.
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;52:377-84.
31. Macfarlane TV, Glenny AM, Worthington HV. Systematic review of population-based epidemiological studies of oro-facial pain. *J Dent* 2001;29:451-67.
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;24:231-9.
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;38:390-6.
34. Genaidy AM, Lemasters GK, Lockey J, et al. An epidemiological appraisal instrument—a tool for evaluation of epidemiological studies. *Ergonomics* 2007;50:920-60.
35. Santaguida P, Raina P. McMaster Quality Assessment Scale of Harms (McHarm) for primary studies. 2008 <http://hiru.mcmaster.ca/epc/mcharm.pdf> (accessed on 27 Oct 2010).
36. Cho MK, Bero LA. Instruments for assessing the quality of drug studies published in the medical literature. *JAMA* 1994;272:101-4.
37. Avis M. Reading research critically. II. An introduction to appraisal: assessing the evidence. *J Clin Nurs* 1994;3:271-7.
38. Cameron I, Crotty M, Currie C, et al. Geriatric rehabilitation following fractures in older people: a systematic review. *Health Technol Assess* 2000;4:111.
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;31:1071-3.
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;39:841-52.
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;292:810-2.
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;271:1615-19.
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. *Fed Focus* (serial online) 2004 <http://www.fedfocus.org/science/london-principles.html> (accessed on 5 Nov 2010).
50. Margetts BM, Vorster V. Evidence-based nutrition—review of nutritional epidemiological studies. *South Afr J Clin Nutr* 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;1:73-7.
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.
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 Agüero 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;16:785-97.
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;288:728-37.
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;19:170-6.
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;94:1551-73.
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;24:687-95.
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;28:690-8.
67. Kreulen CM, Creugers NH, Meijering AC. Meta-analysis of anterior veneer restorations in clinical studies. *J Dent* 1998;26:345-53.
68. Krogh CL. A checklist system for critical review of medical literature. *Med Educ* 1985;19:392-5.
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;80:174-83.
70. Longnecker MP, Berlin JA, Orza MJ, et al. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 1988;260:652-6.
71. Manchikanti L, Singh V, Vilims BD, et al. Medial branch neurotomy in management of chronic spinal pain: systematic review of the evidence. *Pain Phys* 2002;5:405-18.
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;17:512-20.
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;21:503-15.
74. Reisch JS, Tyson JE, Mize SG. Aid to the evaluation of therapeutic studies. *Pediatrics* 1989;84:815-27.
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;57:433-42.
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;152:1289-95.
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;87:193-8.
79. Cowley DE. Prostheses for primary total hip replacement. A critical appraisal of the literature. *Int J Technol Assess Health Care* 1995;11:770-8.
80. Effective Public Health Practice Project. Quality Assessment Tools for Quantitative Studies. 2003.



81. School of Population Health. *EPIQ (Effective Practice, Informatics, and Quality Improvement)*. Faculty of Medical and Health Sciences. University of Auckland, 2004.
82. Fowkes FG, Fulton PM. Critical appraisal of published research: introductory guidelines. *BMJ* 1991;302:1136–40.
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.
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, Eknayan 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;13:920–9.
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;63:502–12.
89. Chokshi DA, Avorn J, Kesselheim AS. Designing comparative effectiveness research on prescription drugs: lessons from the clinical trial literature. *Health Aff (Millwood)* 2010;29:1842–8.
90. Psaty BM, Weiss NS. NSAID trials and the choice of comparators—questions of public health importance. *N Engl J Med* 2007;356:328–30.
91. Strom B. *Pharmacoepidemiology*. 4th edn. Hoboken, New Jersey: 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:S114–20.
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;25:603–5.
94. Conway PH, Clancy C. Charting a path from comparative effectiveness funding to improved patient-centered health care. *JAMA* 2010;303:985–6.
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;363:2228–32.
98. Miettinen OS, Caro JJ. Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. *J Clin Epidemiol* 1989;42:325–31.
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;42:1179–84.
100. Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858–68.

APPENDIX: FLOW DIAGRAM

