PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	How well can we assess the validity of non-randomized studies of medications? A systematic review of assessment tools by the Working Group of the ISPE Comparative Effectiveness Research Special Interest Group
AUTHORS	D'Andrea, Elvira; Vinals, Lydia; Patorno, Elisabetta; Franklin, Jessica; Bennett, Dimitri; Largent, Joan; Moga, Daniela; Yuan, Hongbo; Wen, Xuerong; Zullo, Andrew R; Debray, Thomas; Sarri, Grammati

VERSION 1 – REVIEW

REVIEWER	Prof. Andrea Berghold Medical University of Graz, Institute for Medical Informatics, Statistics and Documentation, Austria
REVIEW RETURNED	30-Sep-2020

GENERAL COMMENTS	D'Andrea and colleagues present a systematic review of tools to assess the validity of non-randomized studies (NRS) of medications. The tools were evaluated against a pre-piloted framework based on the methodological domains recommended by the working group of the International Society for Pharmacoepidemiology Comparative Effectiveness Research (CER) Special Interest Group (SIG). It gives a good overview of checklists and scales available for reporting and critical appraisal of NRS and the domains covered by the different instruments. It is a well written and structured manuscript and should be published in this form. I only noted a small mistake in the strengths and limitation section (p.4, line 46): Tools published in English should be replaced by tools not published in English.

REVIEWER	Aoife Healy Staffordshire University, UK
REVIEW RETURNED	07-Oct-2020

GENERAL COMMENTS	General comments A clear rationale for the need to complete the systematic review is provided, the review followed the appropriate methodology and the results are presented clearly. However, for the Delphi survey element of this study more detailed information is required to provide a reader with a thorough understanding of the process and findings. My main concern is that under the Ethics approval section it is stated that "the study did not involve primary data collection"; while a systematic review doesn't require ethical

	approval, the study also involved a Delphi survey for which ethical approval is required. Abstract When describing data sources there is reference to databases and bibliographies, it is unclear to me what the meaning of “experts’ advice” is within this sentence, I suggest revising this sentence to improve clarity. “Of 44 tools reviewed, 48% assess multiple NRS designs, while 27% and 23% assess respectively case-control and cohort studies only.” – please state what the other 2% of tools assessed. I appreciate the challenge in presenting information on both the systematic review and Delphi survey within a limited word count in the abstract, however, the abstract currently focuses for the majority on the systematic review and I suggest it is revised to provide an overview of the entire study, including the Delphi survey. Introduction A rationale for the need for a Delphi study to identify key methodological challenges for NRS of medications should be added to the introduction. Methods Additional information is required in relation to how “suggestions from experts” were involved in locating tools; Who were these experts? Are they within the CER SIG? In what area are they considered experts? “The findings from the online survey” clarification is required here, is this referring to the Delphi survey? Discussion Page 13 lines 7-14: I think this section which provides the rationale for this study by discussing previous systematic reviews in this area should be moved to the introduction section. Table 2 Define the acronym CER “Tools designed for CER” Table S1 I think some of the search lines require revision #20-#25, its look like there is additional OR “(OR #14 OR #15 OR #16)” and missing brackets “AND OR #14 OR #15 OR #16)”. Supplementary file 1 Limited information on the Delphi study is provided: 1) The results on the demographics of the experts are not provided and it is unclear if all the experts who are invited to participants completed the surveys? 2) How many participated in each round of the Delphi survey, did any dropout? 3) A summary of the results from each of the Rounds should be provided, providing an Supplementary file 2

	It is stated that the Delphi survey consisted of 2 rounds, but I think only 1 round is presented in this file, copies of both rounds should be included in this file.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

D'Andrea and colleagues present a systematic review of tools to assess the validity of non-randomized studies (NRS) of medications. The tools were evaluated against a pre-piloted framework based on the methodological domains recommended by the working group of the International Society for Pharmacoepidemiology Comparative Effectiveness Research (CER) Special Interest Group (SIG).
It gives a good overview of checklists and scales available for reporting and critical appraisal of NRS and the domains covered by the different instruments. It is a well written and structured manuscript and should be published in this form.
I only noted a small mistake in the strengths and limitation section (p.4, line 46): Tools published in English should be replaced by tools not published in English.

RESPONSE: We thank the reviewer for his positive comment. We corrected the clerical error noticed by the Reviewer in the section "Strengths and limitations of this study", and now reads as follows: "Tools not published in English …"

Reviewer: 2

General comments
#1 A clear rationale for the need to complete the systematic review is provided, the review followed the appropriate methodology and the results are presented clearly. However, for the Delphi survey element of this study more detailed information is required to provide a reader with a thorough understanding of the process and findings. My main concern is that under the Ethics approval section it is stated that "the study did not involve primary data collection"; while a systematic review doesn't require ethical approval, the study also involved a Delphi survey for which ethical approval is required.

RESPONSE: We thank the Reviewer for the thoughtful comments. We agree with the Reviewer that part of the study is related to a research survey that involved primary data collection. However, the research imposes no risks of harm and participants in the Delphi survey are competent adults who have been informed through an online meeting that the survey questions were part of a research project and that answering the questions was voluntary. For survey research that does not impose risks on participants and that only enrolls competent adults the value of ethical approval is not required and "lack of ethical oversight should not prevent the manuscript describing the findings from being published," as reported in the peer-reviewed article from Whicher & Wu (reference: Whicher D, Wu AW. Ethics Review of Survey Research: A Mandatory Requirement for Publication? Patient. 2015 Dec;8(6):477-82. doi: 10.1007/s40271-015-0141-0).
In the "Ethical approval" section of the revised manuscript we better specified: "Since the systematic review did not involve primary data collection, and the consensus survey did not impose risks on participants, enrolling only competent adults, the protocol was not submitted for Institutional Review Board approval."

Abstract
#2 When describing data sources there is reference to databases and bibliographies, it is unclear to

me what the meaning of “experts’ advice” is within this sentence, I suggest revising this sentence to improve clarity.

RESPONSE: “Expert’s advice” referred to the list of websites of influential organizations (initiatives, programs, and institutions) suggested by the experts in Pharmacoepidemiology and/or Healthcare Outcome Research who are part of the Comparative Effectiveness Research Special Interest Group (CER SIG; https://www.pharmacoepi.org/communities/sigs/cer/) and participants in the study project. We revised the sentence of the Data sources section of the Abstract as follows: “We systematically searched Pubmed, Embase, Google®, bibliographies of reviews and websites of influential organizations from inception to November 2019.”

#3 “Of 44 tools reviewed, 48% assess multiple NRS designs, while 27% and 23% assess respectively case-control and cohort studies only.” – please state what the other 2% of tools assessed.

RESPONSE: We thank the Reviewer for noticing this discrepancy. We revised the sentence of the Results section of the Abstract as follows: “Of 44 tools reviewed, 48% (n = 21) assess multiple NRS designs, while other tools specifically addressed case-control (n = 12, 27%) or cohort studies (n = 11, 25%) only.”

#4 I appreciate the challenge in presenting information on both the systematic review and Delphi survey within a limited word count in the abstract, however, the abstract currently focuses for the majority on the systematic review and I suggest it is revised to provide an overview of the entire study, including the Delphi survey.

RESPONSE: We have now revised the Abstract by providing more information on the Delphi survey in the Data sources section: “In parallel, we conducted a Delphi survey among the working group of the International Society for Pharmacoepidemiology (ISPE) Comparative Effectiveness Research (CER) Special Interest Group (SIG) to identify key methodological challenges (domains and subdomains) for NRS of medications. Domains and subdomains indicated by the Delphi respondents were employed to evaluate the selected NRS tools.”
We also added in the Results section of the Abstract: “Response rate to the Delphi survey was 73% (35 out of 48 content experts) and a consensus was reached in only two rounds.”

Introduction
#5 A rationale for the need for a Delphi study to identify key methodological challenges for NRS of medications should be added to the introduction.

RESPONSE: We added a rationale in the revised manuscript, paragraph 3 of the Introduction: “There is no agreement on an assessment framework for NRS of pharmacological interventions. Thus, we performed a Delphi survey among international experts in the field of pharmacoepidemiology and health outcome research in order to build consensus for the methodological challenges that may threaten the validity of NRS of medications and that should be evaluated by assessment NRS tools. The main objective of this study was to determine whether the retrieved NRS tools sufficiently address the main methodological challenges recommended by the experts.”

Methods
#6 Additional information is required in relation to how “suggestions from experts” were involved in locating tools; Who were these experts? Are they within the CER SIG? In what area are they considered experts?

RESPONSE: We have now removed this sentence and provided more clarity on the systematic approach of our search. Please also see our response to question #2. The previous sentence

"suggestions from experts" referred to the advice provided by members of the CER SIG on the additional sources to the search of the grey literature (websites of influential organizations, Table S2 in the Supplement 1).

#7 "The findings from the online survey" clarification is required here, is this referring to the Delphi survey?

RESPONSE: For clarity, in the revised version of the manuscript we replaced "online survey" with "Delphi Survey" (line 3 in the section "Data Synthesis" of the Methods).

Discussion
#8 Page 13 lines 7-14: I think this section which provides the rationale for this study by discussing previous systematic reviews in this area should be moved to the introduction section.

RESPONSE: We thank the Reviewer for the comment. We added in the Introduction section of the revised manuscript, 3rd paragraph: "Previously published systematic reviews on assessment tools for NRS were mostly descriptive and did not provide a critical evaluation of the tools content or investigated only a specific type of bias. One systematic review, now outdated, focused only on safety outcomes. Therefore, we conducted a systematic review to identify all existing NRS tools. There is no agreement on an assessment framework for NRS of pharmacological interventions. Thus, we performed a Delphi survey in order to build consensus among international experts in the field of pharmacoepidemiology and health outcome research for the methodological challenges that may threaten the validity of NRS of medications."
Additional details on those systematic reviews are still in the Discussion section (paragraph "Strengths and limitations" of the revised manuscript) to highlight what our study adds to the current literature compared to systematic reviews previously published on similar topics.

Table 2 (the Reviewer was actually referring to Table 3)
#9 Define the acronym CER "Tools designed for CER"

RESPONSE: We added "CER = Comparative Effectiveness research" in Table 3.

Table S1
#10 I think some of the search lines require revision #20-#25, its look like there is additional OR "(OR #14 OR #15 OR #16)" and missing brackets "AND OR #14 OR #15 OR #16)".

RESPONSE: We corrected the typos noted by the Reviewer in the strings #20 to #25 of the Table S1 and corrected the as follows: "(OR #14 OR #15 OR #16)" and "… AND (#14 OR #15 OR #16) …".

Supplementary file 1
Limited information on the Delphi study is provided:
1) The results on the demographics of the experts are not provided and it is unclear if all the experts who are invited to participants completed the surveys?
2) How many participated in each round of the Delphi survey, did any dropout?
3) A summary of the results from each of the Rounds should be provided, providing an

RESPONSE: We provided additional information on the Delphi survey as recommended by the Reviewer.
Specifically, we added in the Supplement 1, lines 32-52: "Thirty-five CER SIG experts participated to the first round of the survey (response rate = 73%). Eighteen respondents worked in academia (51%), 16 respondents worked in industry (46%), and only one was affiliated to a governmental institution (3%). The majority of respondents were from U.S. (n = 26, 74%), the others were from Europe (n = 7,

20%) or from Canada (n = 2, 6%). After participants ranked the relevance of the items and provided comments to the first questionnaire, we reviewed the responses and adjusted the survey to reflect the feedback. Overall, participants considered the presented domains important or extremely important. All domains, other than the domain n. 5 ("Lack of appropriateness of statistical analyses"), were rated relevant with unanimous consent (100% of participants rated them 4 or 5). For the domain n. 5 ("Lack of appropriateness of statistical analyses (with specific mention of adjustment for causal intermediaries, and/or incorrect outcome model specification)" in the Supplement 2), only three participants did not consider it important. Based on individual experts' suggestions, we added the subdomain "External validity of target population" to domain n. 1 "Methods of selecting participants" to address the external validity or generalizability of the study population to the eligible or target population. We removed "Recall bias" from the critical elements of domain n. 3 since this was the only domain to be rated not relevant by most of participants (n = 26, 74%). The reason rationale provided was that, since real world data (RWD) are often collected for intents unrelated to pursue research goals (e.g., administrative purposes), biases such as recall bias, interviewer bias, nonresponse bias are usually reduced or even eliminated [Sørensen HT et al.]. A revised survey was returned to the same participants for reassessment of the content. The response rate to the second round was 89% (n = 31/35)."

Supplementary file 2
It is stated that the Delphi survey consisted of 2 rounds, but I think only 1 round is presented in this file, copies of both rounds should be included in this file.

RESPONSE: We updated the Supplement file 2 and specified which subdomains were added or removed in the questionnaire sent in the second round of the Delphi survey. Specifically, for the subdomain "2.3 Methods of selecting participants*" we specified as follows: "* question added to the questionnaire of the second round of the Delphi survey based on the answers provided on the first round." For the subdomain "3.9 Recall bias#" we reported: "# question removed to the questionnaire of the second round of the Delphi survey based on the answers provided on the first round."