ELSEVIER

Journal of Clinical Epidemiology ■ (2015) ■

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

A principal component analysis is conducted for a case series quality appraisal checklist

Bing Guo[a,*], Carmen Moga[a], Christa Harstall[a], Don Schopflocher[b]

[a]*Health Technology Assessment Program, Institute of Health Economics, #1200 10405 Jasper Avenue, Edmonton, Alberta T5J 3N4, Canada*
[b]*School of Public Health & Faculty of Nursing, University of Alberta, 3-291 Edmonton Clinic Health Academy, 11405-87 Avenue, Edmonton, Alberta T6G 1C9, Canada*

Accepted 28 July 2015; Published online xxxx

Abstract

Objective: Because of a lack of a control group, a case-series study is considered one of the weaker study designs from which to obtain evidence on treatment effectiveness. Under certain circumstances, however, this is the only available evidence to inform health-care decisions. This study's intent was to develop and validate a quality appraisal checklist specifically for case-series studies.

Study design and setting: A modified Delphi process was used to develop the checklist, which was then used by multiple researchers to appraise a random sample of 105 case-series studies. A principal component analysis of these appraisals was conducted to further refine the checklist.

Results: The modified Delphi process resulted in a 20-criterion checklist. The principal component analysis of the appraisals for the 105 case-series studies revealed two components. The first component (10 criteria) indicated the extent to which a case series presented traditional features of a statistical hypothesis-testing paradigm. The second component (seven criteria) indicated whether detailed descriptions of the subjects' characteristics that might feature in the experimental design were present, particularly in judgments about the likelihood of confounding.

Conclusion: This quality appraisal checklist may be useful in assessing case-series studies, but further validation of the checklist is required. © 2015 Elsevier Inc. All rights reserved.

Keywords: Case series; Quality appraisal; Delphi; Validation; Research evidence; Treatment intervention

1. Introduction

Assessment of the safety and efficacy of a treatment intervention requires research evidence from clinical studies. Well-developed methodologies for scrutinizing flaws in design, conduct, analyses, and reporting exist for judging the potential for bias in randomized control trials (RCTs) [1]. In recent years, observational studies (including case-series studies) have been increasingly included in health technology assessments (HTAs) and systematic reviews, but the methodology specifically for assessing case series is less developed [2].

A case-series study is an uncontrolled study that describes a single group of patients who receives a particular intervention and is followed (either prospectively or retrospectively) to observe their outcomes; such a study does not allow for a direct, statistical comparison to a control sample [3,4]. For this reason, a case-series study occupies a low position in the hierarchy of evidence and is considered one of the weaker study designs from which to obtain research evidence on treatment effectiveness [5]. Nevertheless, there are some circumstances where evidence from case-series studies is considered important, necessary, and generally accepted by clinicians [6] and is used in HTAs and systematic reviews for assessing treatment interventions [5]. For example, when examining long-term outcomes, particularly harms of treatment interventions, case-series studies often comprise a large part of the evidence base [7]. Sometimes, case-series studies are the only source of research evidence available to inform decisions about the implementation of new technologies [8–10]. For example, many surgical procedures, therapeutic devices, or rapidly developing interventions are used for conditions where randomization of patients is problematic,

Funding: The project was made possible through a general Knowledge Transfer grant from Alberta Health, Canada. The views expressed herein do not necessarily represent the official policy of Alberta Health.

* Corresponding author: Tel.: 1 780 448 4881; fax: 1 780 448 0018.
E-mail address: bguo@ihe.ca (B. Guo).

http://dx.doi.org/10.1016/j.jclinepi.2015.07.010
0895-4356/© 2015 Elsevier Inc. All rights reserved.

ARTICLE IN PRESS

What is new?

Key findings:

- A principal component analysis of a 20-criterion quality appraisal checklist developed specifically for case series studies revealed two components, with the first component (10 criteria) being related to hypothesis-testing paradigm and the second component (7 criteria) being related to detailed descriptions of subject/intervention characteristics.
- On the basis of the preliminary principal component analysis, no criterion can be eliminated from the 20-criterion checklist.

What this adds to what was known?

- Although a number of quality appraisal checklists for case-series studies were found in the literature, there is a lack of information about whether these tools are formally validated.
- The preliminary validation process serves as a starting point for future large-scale studies in this area.

What is the implication and what should change now?

- We invite health technology assessment researchers/systematic reviewers to use this checklist in assessing evidence from case-series studies and provide feedback to help further refine the checklist.

unethical, or impossible [4,11]. In some cases, the treatment intervention of interest is the last resort for the condition; thus, even a nonrandomized comparison (such as in a cohort study) would not be possible.

Because case-series studies are prone to various types of biases in relation to selection, performance, detection, attrition, and reporting, the derived results are generally ranked as low quality [12]. Assessing the robustness of the evidence of uncontrolled studies remains challenging [10]. In contrast to the availability of well-developed quality appraisal tools for RCTs (eg, the Cochrane Collaboration risk of bias tool [1]) or for non-RCTs (eg, A Cochrane Risk Of Bias Assessment Tool for Non-Randomized Studies of Interventions [13]) in which two patient groups are compared, few quality appraisal tools exist for case-series studies that have only one group of patients.

Although some criteria in nonspecific checklists can be common between case-series studies and other types of studies, using nonspecific checklists for case-series studies represents a redundant practice in the field of HTA and systematic reviews, as these checklists usually include many criteria not relevant to case-series studies (eg, criteria related to the control group). Some quality appraisal tools have been developed specifically for case series [14] or have been used for assessing case-series studies [15,16]. However, no widely accepted and validated quality appraisal tool exists for assessing the methodologic quality of case-series studies [2,17]. Furthermore, there is no consensus about which criteria should be included for this purpose [12]. As quality appraisal tools should be rigorously developed, evidence based, valid, reliable, and easy to use [2], there has been an increased interest in developing such a quality appraisal checklist.

The objective of this study was to validate a quality appraisal checklist for case-series studies that was developed through a modified Delphi process.

2. Methods

The development and validation of the checklist consisted of three separate steps: the Delphi process, the pilot exercise, and the validation process. As detailed descriptions of the first two steps are available in a previous report [18], this article focuses mainly on the validation process, with only a brief summary of the first two steps for context.

2.1. The Delphi process

The Delphi process, a structured communication process characterized by purposive sampling of experts in the field of interest, panelist anonymity, multiple iterations, and feedback of statistical aggregation of group responses [19,20], was used for final consensus of the criteria.

A group of seven experienced HTA researchers (five from the Institute of Health Economics [IHE], Canada, one from ASERNIP-S, Australia, and one from the AETS, Spain), participated in the modified four-round e-mail–based Delphi process to develop the quality appraisal checklist. Each panelist received an initial list of 30 criteria (Table 1) compiled from a literature search [12,21–24] and were asked to rate the relevance and importance of each criterion based on a five-point Likert scale that ranged from 1 (very important) to 5 (not important at all). A criterion was included in or excluded from the final checklist if at least five of the seven panelists (ie, 70% agreement among panelists, a cutoff decided a priori) judged that criterion either very important (rank 1) or not important at all (rank 5). The panelists also had the opportunity to suggest new criteria and how to improve clarity of the final checklist.

Each panelist received a personalized summary of his/her own rating results (blind to other individuals' rating to ensure anonymity) and the aggregated group ratings (results and distribution of ratings from all panelists), along with any newly suggested criteria. The panelists' responses were analyzed by a biostatistician (D.S.), who was not part of the Delphi panel and was also blinded as to the identities of the panelists.

ARTICLE IN PRESS

Table 1. Thirty criteria included in the initial list

Initial criterion
1. Is the hypothesis/aim/objective of the study stated in the abstract, introduction, or methods section?
2. Are the characteristics of the patients included in the study clearly described?
3. Were the case series collected in more than one center?
4. Are the eligibility criteria (inclusion and exclusion criteria) explicit and appropriate?
5. Were data collected prospectively?
6. Were patients recruited consecutively?
7. Did patients enter the study at a similar point in the disease?
8. Were the subjects recruited during the same period of time?
9. Did the authors describe the intervention?
10. In addition to intervention, did the patients receive any cointervention?
11. Was there loss to follow-up reported?
12. Are outcomes (primary, secondary) clearly defined in the introduction or methodology section?
13. Were the outcomes assessed blind/independent to intervention status?
14. Did the authors use accurate (standard, valid, reliable), objective methods to measure the outcomes?
15. Did the authors use standardized subjective measures (eg, questionnaires or patient symptoms interview forms) to assess the outcomes?
16. Were outcomes assessed before and after intervention?
17. Was the length of follow-up clearly described/reported?
18. Were the statistical tests used to assess the primary outcomes appropriate?
19. Have actual probability values been reported (eg, $P = 0.035$ rather than $P < 0.05$) for the primary outcome measurements except where the probability value is less than 0.001?
20. Does the study provide estimates of the random variability in the data for the primary outcomes (eg, standard error, standard deviation, confidence intervals)?
21. Was there a discussion/assessment of possible confounders?
22. Was the analysis of outcomes based on the number of patients available at the time when the follow-up measures were taken?
23. Are the main findings of the study clearly described?
24. Are outcomes of the study stratified (eg, based on follow-up periods, etiologies, cointervention)?
25. Do the study's findings respond to research objective(s)/question(s)?
26. Are adverse events that may be a consequence of the intervention reported?
27. Are results based on data dredging?
28. Are the conclusions of the study supported by results?
29. Are the limitations of the study taken into consideration?
30. Is there a competing interest statement about the type and source of support received for the study or about the relationship of the author(s) or other contributors with the manufacturer of the technology?

2.2. The pilot exercise

In response to a suggestion from most of the panelists mentioned in Section 2.1, two researchers (C.M. and B.G.) developed draft guidance to help standardize the rating for each criterion. The resulting quality appraisal checklist from the Delphi process along with the draft guidance was used by three researchers to appraise the quality of 13 case-series studies included in an HTA project on the safety and effectiveness of islet transplantation for the treatment of type 1 diabetes [18]. Before starting the quality appraisal process, two of the three researchers thoroughly discussed the clinical conditions and treatment interventions and adapted the guidance document accordingly. Interrater reliability was measured by the Kappa statistic, and mean and standard deviations were calculated for the ratings for each study across the three reviewers. In addition, the draft checklist and guidance were used by researchers from other organizations.

2.3. The validation process

The small number of raters and case-series studies assessed in the pilot made it difficult to obtain reliable and stable estimates of the measurement properties of the checklist, leading to the initiation of a formal validation process. Because face and content validity of the checklist were evaluated via the Delphi panelists who judged the relevance and representativeness of its criteria, the validation process focused only on the relationships between the items in the evaluation of case-series studies. The objective of the validation process was to provide information on the utility of the checklist that can contribute to judgments about its validity.

One hundred five case-series studies were randomly selected by two reviewers (C.M. and B.G.) from a pool of 13,138 citations obtained by a literature search for case-series studies published in English between 2006 and 2012. The determination of the sample size of 105 studies was based on the suggested guidelines of 10 times the number of responses to number of items (210 sets of evaluator responses on the 20 individual items), a common rule of thumb for the minimum ratio of responses to items for principal component/exploratory factor analyses [25]. Most of the 105 articles focused on the effectiveness of different surgeries (eg, orthopedic, ophthalmologic, neurologic, gastrointestinal, cardiovascular, or obstetrical surgeries), whereas some studies focused on nonsurgical treatments for different conditions (eg, eye diseases, malignancies, mental health disorders, posttransplant, or poststroke).

ARTICLE IN PRESS

Six HTA researchers, four of whom also participated in the Delphi process, used the 20-criterion checklist and the draft guidance to appraise the quality of an assigned set of the 105 studies. Each of the six researchers independently assessed 35 studies in such a way that each possible pairing of researchers assessed seven studies in common. This design aimed to reduce researchers' workload while still allowing calculation of meaningful statistical properties of the checklist. Researchers had the opportunity to use a predeveloped feedback form to provide their comments on the feasibility, relevance, and understandability of the criteria/guidance, any difficulties with the rating of the criteria, and suggestions for new criteria or improving the guidance. Researchers also recorded the time they spent appraising each study.

An exploratory principal component analysis [25,26] was conducted on these ratings to examine the interrelationships among the criteria and identify clusters of criteria that share considerable variation and to suggest potential summary scoring of the checklist. Principal component analysis is typically used as a method of reducing the number of variables required to describe a set of scores, while maximizing the amount of information retained. It operates by combining the scores of the original variables into components and is guided by the correlations of the items with each other [26]; typically, variables which correlate highly with each other will be combined into a single component. Principal component analysis is related to factor analysis, which is a complex variant typically used to locate/postulate underlying unities which "cause" sets of items to be answered similarly [27,28].

Interrater reliability was measured by examining the correlations in the scores between the six researchers.

3. Results

3.1. Quality appraisal checklist

A 100% response rate was reached in the first three rounds of the Delphi process, whereby an 18-criterion checklist was developed from the initial 30 criteria (Table 1). Two criteria (prospective study design and blind assessment of outcomes) were added to the final checklist on the basis of a supplemental literature search, resulting in the final 20-criterion quality appraisal checklist (Table 2). The full checklist along with detailed guidance is presented in Appendix A at www.jclinepi.com. The checklist examines the main domains of a case-series study

Table 2. IHE QA checklist form for case-series studies

Criterion	Rating		
	Yes	Partial/unclear	No
Study objective			
1. Was the hypothesis/aim/objective of the study clearly stated?			
Study design			
2. Was the study conducted prospectively?			
3. Were the cases collected in more than one center?			
4. Were patients recruited consecutively?			
Study population			
5. Were the characteristics of the patients included in the study described?			
6. Were the eligibility criteria (i.e. inclusion and exclusion criteria) for entry into the study clearly stated?			
7. Did patients enter the study at a similar point in the disease?			
Intervention and cointervention			
8. Was the intervention of interest clearly described?			
9. Were additional interventions (cointerventions) clearly described?			
Outcome measures			
10. Were relevant outcome measures established a priori?			
11. Were outcome assessors blinded to the intervention that patients received?			
12. Were the relevant outcomes measured using appropriate objective/subjective methods?			
13. Were the relevant outcome measures made before and after the intervention?			
Statistical analysis			
14. Were the statistical tests used to assess the relevant outcomes appropriate?			
Results and conclusions			
15. Was follow-up long enough for important events and outcomes to occur?			
16. Were losses to follow-up reported?			
17. Did the study provided estimates of random variability in the data analysis of relevant outcomes?			
18. Were the adverse events reported?			
19. Were the conclusions of the study supported by the results?			
Competing interests and sources of support			
20. Were both competing interests and sources of support for the study reported?			

Abbreviation: IHE, Institute of Health Economics; QA, quality appraisal.

including study design, population, intervention and cointervention(s), outcome measures, statistical analysis, results and conclusions, and competing interest and sources of financial support. All panelists agreed that this checklist should be used for treatment intervention case-series studies only.

3.2. Feedback from the pilot exercise

Interrater reliability analysis showed that researchers who discussed and adapted the checklist before conducting the quality appraisal had better agreement in their ratings. Researchers reported the total quality rating scores (1 point for a ''yes'' answer, 0 for an ''unclear'' or ''no'' answer) for each study but did not use any cutoff value as inclusion or exclusion criteria. The methodologic quality of included studies was discussed in the report by highlighting ratings of five criteria (prospective, consecutive recruitment, multicenter study, before and after measurement, and loss to follow-up) that were thought to be most important and relevant to the research topic.

In general, researchers found the checklist easy to use and helpful in identifying a study's strength and limitations. The process of discussing relevant clinical aspects and adapting the guidance to the topic in the pilot phase helped improve the clarity and feasibility of the checklist and the guidance.

3.3. Empirical analysis of the checklist

3.3.1. Time required for quality appraisal

The reported time for appraising each study ranged from 5 to 110 minutes (median 15, interquartile range [IQR]: 29:12); more time was required when the information was not clearly presented or additional referenced background material had to be retrieved. Researchers generally felt more comfortable with the checklist after completing the first few studies and found a reduction in the time required to appraise the next set of studies.

3.3.2. Frequency of responses

As summarized in Table 3, within the 105 case-series studies, 10 criteria (nos. 1, 5, 8, 10, 12, 13, 15, 16, 18, and 19) were generally present and well reported; seven criteria (nos. 4, 6, 7, 11, 14, 17, and 20) were generally present but were reported only partially or were unclear or vague. The three remaining criteria (nos. 2, 3, and 9) were generally not present in the studies.

3.3.3. Principal component analysis

All 210 ratings reported by the six researchers were entered into a principal component analysis of the 20-criterion checklist. The largest two components accounted for 23.9% of the variance. As shown in Fig. 1, the first of these two components brings together 10 criteria which are closely associated (the circle on the lower right corner),

Table 3. Frequency of responses in 105 case-series studies

Number	Criterion	Responses[a]		
		Yes (%)	Partial/unclear (%)	No (%)
Generally present				
1	Hypothesis/aim/objective	74.3	22.9	2.9
5	Characteristics described	76.2	17.6	6.2
8	Intervention clearly described	84.3	12.4	3.3
10	Outcome measures established a priori	83.8	7.6	8.6
12	Outcomes measured appropriately	77.1	11.0	11.9
13	Outcomes measured before–after	72.4	10.0	17.6
15	Length of follow-up reported	79.0	14.8	6.2
16	Loss to follow-up reported	73.8	7.6	18.6
18	Adverse events reported	70.0	6.7	23.3
19	Conclusions supported by results	71.4	23.8	4.8
Generally partially presented or unclear				
4	Consecutive recruitment	56.2	31.9	11.9
6	Inclusion/exclusion criteria	52.9	37.1	10.0
7	Similar point of entry in study	45.7	35.7	18.6
11	Outcomes assessed blinded	14.8	53.8	31.4
14	Statistical tests appropriate	59.5	21.0	19.5
17	Estimates of random variability in data analysis or relevant outcomes	50.0	11.0	39.0
20	Competing interest and sources of support reported	35.2	42.9	21.9
Generally not present				
2	Study conducted prospectively	26.2	18.6	55.2
3	Multicenter study	11.4	23.8	64.8
9	Cointerventions reported	27.1	27.1	45.7

[a] One hundred five case-series studies; each of the six researchers reviewed 35 studies, two researchers reviewed the same seven studies.

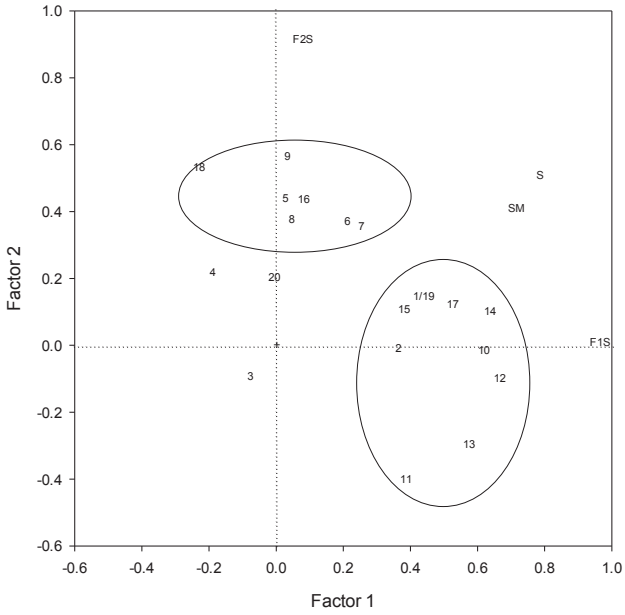


Fig. 1. Scoring the 20-criterion checklist.

whereas the second brings together seven criteria (the circle on the upper left part). Three criteria (multicenter study, consecutive recruitment, and competing interest and sources of support) do not correlate highly with either of the two component groups. Consecutive recruitment (criterion 4) is a key component of a case-series study to minimize selection bias but here falls out of the analysis in the ''Other'' category. This is because this criterion did not get answered over the six researchers and 105 studies in a manner broadly consistent with either of the two groupings of items that do tend to show similarity of response over items and researchers.

An extension analysis of the composite variables (F1S and F2S), which were obtained by summing the two groups of circled criteria, showed that the composite variables have higher loadings and are univocal (ie, correlating highly with only one factor). This suggests that these combinations of items can be used in further analyses as partial summaries of the 20 criteria. No criterion can be eliminated from the 20-criterion checklist on the basis of the preliminary principal component analysis.

Table 4 illustrates that the first component appears to bring together the traditional features of the execution of a statistical hypothesis-testing paradigm. The second component appears to bring together more detailed descriptions of the subjects' characteristics that might feature in the experimental design, particularly in judgments about the likelihood of confounding.

3.4. Interrater reliability

Analysis of correlations between two researchers was based only on the seven studies rated in common, which makes correlations unstable. The interrater reliability results are not reported because of the very small number of studies rated by two researchers.

4. Discussion

4.1. Strength and limitations

The researchers involved in the Delphi and validation processes were a homogeneous group with an intimate knowledge of the various critical appraisal tools used in the field of HTA, as well as the values and limitations of using case-series studies for assessing the safety and effectiveness of treatment interventions. Although no clinical research experts were involved in the Delphi process, some panelists in this group had medical backgrounds and clinical research experience. The involvement of an experienced biostatistician helped to ensure the appropriateness of the study design, data analysis, and interpretation of results.

One of the limitations of this study is the small number (seven) of Delphi panelists; in general, 10 to 15 panelists are considered sufficient for a Delphi study that is focused and the background of the Delphi participants is homogenous [29]. Limited resources prevented us from recruiting more researchers from a broader HTA community. It is hoped that funding may be available in the future, and that other HTA researchers will be interested to continue this work.

Table 4. Preliminary results of the principal component analysis

Component 1	Component 2	Other
1. Hypothesis/aim/objective	5. Characteristics described	3. Multicenter study
2. Study conducted prospectively	6. Inclusion/exclusion criteria	4. Consecutive recruitment
10. Outcome measures established a priori	7. Similar disease points of entry in study	20. Competing interest and sources of support
11. Outcomes measured blinded	8. Intervention clearly described	
12. Outcomes measured appropriately	9. Cointerventions reported	
13. Outcomes measured before–after	16. Loss to follow-up reported	
14. Statistical tests appropriate	18. Adverse events reported	
15. Length of follow-up reported		
17. Estimates of random variability in data analysis of relevant outcomes		
19. Conclusions supported by results		

4.2. Comparison to other quality appraisal checklists

The literature search conducted for this project revealed a paucity of checklists specifically designed for case-series studies, as well as a lack of reported details on the development of such checklists. One specific tool of interest was the 13-criterion quality appraisal checklist developed by Yang [14] for case-series studies in the field of herbal medicine. Although a Delphi process was used in this study, the initial selection of criteria for this checklist was based on expert opinion. Compared to our checklist, Yang et al.'s checklist has fewer criteria, although, in some cases, one single criterion covered more than one aspect and overlapped with those on our checklist. Six criteria are common to both checklists: clear aim of the study, explicit/clear patient inclusion and exclusion criteria, clear description of the treatment intervention, objective outcome assessment, appropriate data/statistical analysis, and reporting of adverse events. Criteria regarding prospective study design, multicenter study, consecutive patient recruitment, the rate and length of follow-up, conclusion supported by results, and competing interest/source of support are incorporated only in our checklist. In addition, several criteria are included in the checklist by Yang et al. but not in ours, such as appropriate study design, description of disease/conditions being treated, clear rationale for the treatment protocol, adequate description of methods/procedures, clear report of all outcome measures, and relevant and complete data collection. The specific focus on herbal medicine in Yang et al.'s study may have contributed to the selection of some different criteria by the experts in that field. However, no guidance was provided to rate the 13 criteria of Yang et al.'s checklist, which could make it difficult for users to be consistent in their ratings and their understanding of the criteria.

We also found several other shorter quality appraisal checklists for case-series studies, such as those reported in the Center for Reviews and Dissemination's Guidance, published in 2001 [15], and in other studies by Young et al. [30], Chan and Bhandari [31], and Vere and Joshi [32]. The number of included criteria in these studies ranged from three to 10. However, no published validation studies of these checklists were identified through our literature search.

Our quality appraisal checklist, although contains 20 criteria, is easy to use with the clearly defined guidance. In most cases, less than 30 minutes was required for appraising each study, indicating its user-friendly nature.

4.3. How to use the checklist

For our quality appraisal checklist, the term "quality" covers both risk of bias and quality of reporting. "Bias" refers to systematic deviation from the true underlying effect brought about by poor study design or conduct in the collection, analysis, interpretation, publication, or review of data [33]. Although "reporting" is not a direct measure of the quality of a study, it allows a reader to assess the validity and applicability of the study's findings.

In our checklist, several criteria (eg, prospective study, consecutive recruitment, predefined inclusion/exclusion criteria, before and after outcome measurement, and sufficient follow-up) examine how the study was executed, whereas other criteria (eg, clear statement of study objective, description of patient characteristics, intervention and cointerventions, or reporting of adverse events, and competing interest and sources of support) focused on the quality of the reporting. The principal component analysis revealed that many reporting-related criteria were present and well reported in most of the 105 studies, whereas some "quality" criteria were only partially present, unclear, or not present (Table 3), suggesting common shortcomings across case-series studies.

The currently available reporting guidance for observational studies [34] did not specifically address case-series studies, and the recently published case report guidelines were proposed for clinical case reports but not for case-series studies [35]. Thus, our checklist may also serve as a reporting guidance for publishing case-series studies.

Our analysis provided no basis for cutoff scores by which study quality might be distinguished. As empirical evidence does not support such a quality scale [36], no scale or numeric scoring system was developed for our checklist. Our quality appraisal checklist has been used by other researchers in their systematic reviews in such a way that 1 point was given for each "yes" answer (ie, each criterion was weighted equally), for a maximum of 20 points [37,38]. The rational of equal weighting for each criterion in our checklist remains to be determined in future research. According to the current guidance, the quality appraisal results should not be used to exclude studies from further analyses, but should be incorporated in data analysis and synthesis, and interpretation of research findings [33]. As validation of a quality appraisal tool is an ongoing process, feedback from future users of the tool may help determine the useful measurement properties of the checklist and further explore the optimal use of the checklist.

The checklist should be tailored to meet the needs of the research project, taking into account the relevance and importance of hypothesis testing versus description of subject/intervention characteristics. We suggest using the full checklist, although criteria that are not applicable or appropriate for a specific research project can be omitted from the checklist. Interpretation of the quality appraisal results should be based on individual criterion. Some key criteria most relevant and critical to specific reviews can be highlighted when discussing the methodologic quality of included studies. For example, in the aforementioned pilot exercise on islet transplantation, we attempted to highlight five important items (prospective, consecutive recruitment, multicenter study, before and after measurement, and loss

ARTICLE IN PRESS

to follow-up) in our discussion and interpretation of quality assessment results.

Before starting the quality appraisal process, the guidance for the checklist should be discussed and adapted to reflect the most relevant and important clinical aspects. For an important or complex research topic, a multidisciplinary team consisting of an HTA researcher, clinician, epidemiologist, biostatistician, and methodologist should be involved in the discussions, when possible.

5. Conclusions

A collaborative effort of HTA researchers from Canada, Australia, and Spain resulted in a 20-criterion quality appraisal checklist with a clearly defined guidance specifically for case-series studies. The face and content validity of the checklist were evaluated via a modified Delphi process, and the relationships between the items were evaluated by a principal component analysis. No criterion can be deleted from the checklist based on the preliminary analysis. Feedback from IHE researchers, who piloted the checklist and researchers from other organizations who opted to use the checklist, helped to refine the checklist and its guidance.

The IHE quality appraisal checklist accompanied by a user-friendly guidance serves as a good starting point to examine the quality of reporting and risks of bias of a case-series study. With some adaptation, the checklist may be useful for future HTA work.

Acknowledgments

The authors would like to thank Dr. Ann Scott, Mr. Ken Bond, Ms. Liz Dennett, and Ms. Paula Corabian from IHE, Canada; Dr. Maria Ospina from the University of Alberta, Canada; Dr. Alun Cameron from ASERNIP-S, Australia; and Dr. Iñaki Imaz Iglesia from AETS, Institute of Health Carlos III, Spain, for their participation in the Delphi and validation processes.

The authors would also like to thank Ms. Wendy McIndoo from IHE for her assistance with the Delphi process; and especially Dr. Dave Slessor from the UK, Ms. Nienke Hennequin-Hoenderdos from the Netherlands, and Ms. Veronique Young-Chan-Nam and Mr. Daniel Tran from Canada for using the checklist in their projects and for providing their comments and suggestions on the draft checklist.

References

[1] Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.

[2] 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(3):666−76.

[3] Facey K, INAHTA. Health technology assessment (HTA) glossary. 1st ed. Stockholm: INAHTA; 2006. Available at http://www.inahta.org/upload/HTA_resources/Edu_INAHTA_glossary_July_2006_final.pdf. Accessed July 30, 2014.

[4] Marko NF, Weil RJ. The role of observational investigations in comparative effectiveness research. Value Health 2010;13(8):989−97.

[5] Kempen JH. Appropriate use and reporting of uncontrolled case series in the medical literature. Am J Ophthalmol 2011;151:7−10.

[6] Jabs DA. Improving the reporting of clinical case series. Am J Ophthalmol 2005;139:900−5.

[7] Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. Ann Intern Med 2005;142:1090−9.

[8] Stein K, Dalziel K, Garside R, Castelnuovo E, Round A. Association between methodological characteristics and outcome in health technology assessments which included case series. Int J Technol Assess Health Care 2005;21:277−87.

[9] Chambers D, Rodgers M, Woolacott N. Not only randomized controlled trials, but also case series should be considered in systematic reviews of rapidly developing technologies. J Clin Epidemiol 2009;62:1253−60.

[10] Norris SL, Atkins D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions. Ann Intern Med 2005;142:1112−9.

[11] Ray JG. Evidence in upheaval: incorporating observational data into clinical practice. Arch Intern Med 2002;162:249−54.

[12] Dalziel K, Round A, Stein K, Garside R, Castelnuovo E, Payne L. Do the findings of case series studies vary significantly according to methodological characteristics? Health Technol Assess 2005;9: iii-iv, 1-146.

[13] Sterne JAC, Higgins JPT, Reeves BC, editors. A Cochrane Risk Of Bias Assessment Tool: for Non-Randomized Studies of Interventions (ACROBAT-NRSI). Version 1.0.0. 2014. Available at: http://www.riskofbias.info. Accessed May 28, 2015.

[14] Yang AW. Assessing quality of case series studies: development and validation of an instrument by herbal medicine CAM researchers. J Altern Complement Med 2009;15:513−22.

[15] Khan K, Riet G, Glanville J, Sowden A, Kleijnen J, editors. Undertaking systematic reviews of research on effectiveness: CRDs guidance for those caring out or commissioning reviews. 2nd edition. York, UK: NHS Centre for Reviews and Dissemination; 2001.

[16] Jonsson L, Andreasen N, Kilander L, Soininen H, Waldemar G, Nygaard H, et al. Patient-and proxy-reported utility in Alzheimer disease using the EuroQoL. Alzheimer Dis Assoc Disord 2006;20:49−55.

[17] Mallen C, Peat G, Croft P. Quality assessment of observational studies is not commonplace in systematic reviews. J Clin Epidemiol 2006;59:765−9.

[18] Moga C, Guo B, Schopflocher D, Harstall C. Development of a quality appraisal tool for case series studies using a modified Delphi technique. Edmonton: Institute of Health Economics; 2012. Available at http://www.ihe.ca/advanced-search/development-of-a-quality-appraisal-tool-for-case-series-studies-using-a-modified-delphi-technique. Accessed September 14, 2015.

[19] Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. PLos Med 2011;8:e1000393.

[20] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000;32:1008−15.

[21] National Collaborating Center for Acute Care. Preoperative tests: The use of routine preoperative tests for elective surgery: Evidence, methods & guidance 2003. Available at: http://www.nice.org.uk/guidance/cg3/evidence/cg3-preoperative-tests-full-guideline3. Accessed May 28, 2015.

ARTICLE IN PRESS

[22] 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(6):377−84.
[23] West S, King V, Carey TS, Lohr KN, McKoy N, Lux L. Systems to rate the strength of scientific evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality; 2002.
[24] Elwood JM. Critical appraisal of epidemiological studies and clinical trials. 3rd ed. Oxford: Oxford University Press; 2007.
[25] Costello A, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assess Res Eval 2005;10:1−9.
[26] Sainani KL. Introduction to principal components analysis. PM & R 2014;6(3):275−8.
[27] Matsunaga M. How to factor-analyze your data right: do's, don'ts, and how-to's. Int J Psychol Res 2010;3:97−110.
[28] Portney LG, Watkins MP. Multivariate analysis. 3rd ed. Foundations of clinical research: applications to practice Upper Saddle River, NJ: Pearson/Prentis Hall; 2009:685−726.
[29] Hsu C-C, Standord BA. The Delphi technique: making sense of consensus. PARE 2007;12(10):1−8.
[30] Young J, Hyde C, Fry-Smith A, Gold L. Lung volume reduction surgery for chronic obstructive pulmonary disease with underlying severe emphysema. West Midlands Development and Evaluation Committee Report. Birmingham: University of Birmingham; 1999.
[31] Chan K, Bhandari M. Three-minute critical appraisal of a case series article. Indian J Orthop 2011;45:103−4.
[32] Vere J, Joshi R. Quality assessment of prospective case series of dental implant surgery and prosthodontics published between 2004 and 2008: a systematic review. Int J Oral Maxillofac Implants 2012;27:865−74.
[33] Centre for Reviews and Dissemination. Systematic reviews. CRD's guidance for undertaking reviews in health care. Layerthorpe York: York Publishing Services Ltd. 2009.
[34] von Elm E, 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. Lancet 2007;370(9596):1453−7.
[35] Gagnier JJ, Kienle G, Altman DG, Moher D, Sox H, Riley D. The CARE guidelines: consensus-based clinical case reporting guideline development. J Med Case Rep 2013;7:223. 223−7.
[36] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. JAMA 1999;282:1054−60.
[37] Slessor D, Hunter S. To be blunt: are we wasting our time? Emergency department thoracotomy following blunt trauma: a systematic review and meta-analysis. Ann Emerg Med 2015;65:297−307.
[38] Hennequin-Hoenderdos N, Slot, Van der W G. The incidence of complications associated with lip and/or tongue piercings: a systematic review. Int J Dent Hyg 2015;10:1−12.

ARTICLE IN PRESS

Appendix A. IHE's quality appraisal checklist for assessing case-series studies

Table A.1. Quality appraisal checklist for case-series studies and guidance

Criterion and guidance for rating each criterion
Study objective
1. Was the hypothesis/aim/objective of the study clearly stated? Yes: The hypothesis/aim/objective of the study was clearly reported (includes patients, intervention, and outcome). Partial: Only one or two components (patients, intervention, or outcome) were included. No: The hypothesis/aim/objective was not reported.
Study design
2. Was the study conducted prospectively? Yes: It was clearly stated that the study was conducted prospectively. Unclear: Unclear or no information was provided. No: The study clearly stated it was a retrospective study.
3. Were the cases collected in more than one center? Yes: Cases were collected in more than one center (multicenter study). Unclear: Unclear where the patients came from. No: Cases were collected from one center.
4. Were patients recruited consecutively? Yes: There was a clear statement or it was clear from the context that the patients were recruited consecutively; or the study stated that all eligible patients were recruited. Unclear: No information was provided about the method used to recruit patients in the study. No: The study clearly stated that patients were not recruited consecutively; or the patients were recruited based on other criteria such as access to intervention determined by the distance or availability of resources.
Study population
5. Were the characteristics of the patients included in the study described? Yes: All the most relevant characteristics of the patients were reported (eg, number, age, gender, ethnicity, severity of disease/condition, comorbidity, or etiology). Partial: Some, but not all, of the most relevant characteristics were reported. No: Only the number of patients was reported. Note: Assessor(s) should decide which aspects are important before using the checklist.
6. Were the eligibility criteria (ie, inclusion and exclusion criteria) for entry into the study clearly stated? Yes: Both inclusion and exclusion criteria were reported. Partial: Either the inclusion or exclusion criteria were reported. No: Neither inclusion nor exclusion criteria were reported. Note: Assessor(s) should decide which aspects are important before using the checklist.
7. Did patients enter the study at a similar point in the disease? Yes: It was clear from the baseline data presented in the study (eg, tables of patients' characteristics) that most (at least 80%) patients entered the study at a similar point in terms of the duration and severity of the disease/condition and the presence of comorbidities/complications. Unclear: There was no baseline information on patients' characteristics to make a judgment. No: There was a wide range in the severity of the disease/condition and comorbidities/complications in patients at baseline. Note: Assessor(s) should decide which aspects are important before using the checklist. It might be useful to discuss with specialists to determine the most important aspects that should be considered.
Intervention and cointervention
8. Was the intervention of interest clearly described? Yes: All the most relevant characteristics of the intervention were reported (eg, dosage, frequency or duration of intervention, administration methods, technical parameters, or characteristics of a device). Partial: Some, but not all, of the most relevant characteristics were reported. No: Only the name of the intervention was reported. Note: Assessor(s) should decide which aspects are important before using the checklist.
9. Were additional interventions (cointerventions) clearly described? Yes: All the most relevant characteristics of the cointervention(s) were reported (eg, different type, dosage, frequency of administration, or duration); or the study clearly stated that a cointervention was not administered for clinical reasons. Partial: Some, but not all, of the most relevant characteristics of the cointervention were reported. No: No information about cointervention(s) was provided; or only the name(s) of the cointervention(s) were mentioned. Note: Assessor(s) should decide which aspects are important before using the checklist.
Outcome measures
10. Were relevant outcome measures established a priori? Yes: All relevant outcome measures were stated in the introduction or methods section. Partial: Some, but not all, of the relevant outcome measures were stated in the introduction or method section. No: None of the relevant outcome measures were stated in the introduction or method section.

(*Continued*)

ARTICLE IN PRESS

Table A.1. Continued

Criterion and guidance for rating each criterion
11.Were outcome assessors blinded to the intervention that patients received?
Yes: The relevant outcomes were assessed by individuals who were not aware of the intervention. Answer yes when blinding is not applicable or is unnecessary (eg, mortality).
Unclear: The study did not report whether the outcome assessors were aware of the intervention.
No: It was clearly stated or obvious from the context that the relevant outcomes were analyzed by individuals who were aware of the intervention provided to patients.
12.Were the relevant outcomes measured using appropriate objective/subjective methods?
Yes: All relevant outcomes were measured with appropriate methods. These measures can be objective (eg, gold standard tests or standardized clinical tests), subjective (eg, self-administered questionnaires, standardized forms, or patient symptoms interview forms), or both.
Partial: Some, but not all, relevant outcomes were measured with appropriate methods.
No: The methods used to measure the relevant outcomes were inappropriate.
Note: Assessor(s) should decide which methods are appropriate before using the checklist.
13.Were the relevant outcome measures made before and after the intervention?
Yes: The relevant outcome measures were made before and after intervention; or the baseline measurements were not possible (eg, death).
Unclear: The study did not report when the outcome measures were made.
No: The outcome measures were only made after intervention.
Statistical analysis
14.Were the statistical tests used to assess the relevant outcomes appropriate?
Yes: The statistical tests were used appropriately (eg, parametric test for normally distributed population vs. nonparametric test for non-Gaussian population). Answer ''yes'' if no statistical analysis was performed and reasons for this were stated.
Unclear: The statistical tests were not described in the methods section of the study.
No: The statistical tests used were inappropriate.
Note: Assessor(s) should decide which statistical tests are appropriate before using the checklist. Seek expert assistance if necessary.
Results and conclusions
15.Was follow-up long enough for important events and outcomes to occur?
Yes: It was clear from the information provided that the follow-up period was long enough for most (at least 80%) patients, to allow for important events and outcomes (eg, changes in clinical status, adverse events) to occur.
Unclear: The length of follow-up was not clearly reported.
No: It is clear from the information provided that the follow-up period was not long enough to allow for important events and outcomes to occur.
Note: Assessor(s) should define the appropriate duration of follow-up for each outcome of interest (eg, short-term and long-term adverse events).
16.Were losses to follow-up reported?
Yes: The number or proportion of patients lost to follow-up was clearly reported; the authors reported outcome results on all patients initially included; or the number lost to follow-up can be subtracted from the number of patients enrolled and the number of patients included in the final analysis.
Unclear: There was a discrepancy between the number or proportion of patients reported in tables, figures, and text.
No: The number or proportion of patients lost to follow-up was not reported.
17.Did the study provided estimates of random variability in the data analysis of relevant outcomes?
Yes: The estimates of the random variability (eg, standard error, standard deviation, confidence interval for normally distributed data or range, and interquartile range for nonnormally distributed data) were reported for all the relevant outcomes or could be calculated from the raw data presented in the study.
Partial: The estimates of the random variability were reported for some, but not all, of the relevant outcomes.
No: The estimates of the random variability were not reported for any of the relevant outcomes.
18.Were the adverse events reported?
Yes: The undesirable or unwanted events during the study period or within a prespecified time period were reported; or the absence of adverse event(s) was mentioned in the study.
Partial: Some, but not all, important adverse events were reported.
No: There was no statement about the presence or absence of adverse events.
Note: Assessor(s) should decide which adverse events are most important. Seek clinical expert assistance if necessary.
19.Were the conclusions of the study supported by the results?
Yes: The conclusions of the study were supported by the evidence presented in the results and discussion sections.
Unclear: Unclear conclusion statement that makes it difficult to link the presented evidence to conclusions.
No: The conclusions were not supported by the evidence presented in the results and discussion sections.
Competing interests and sources of support
20.Were both competing interests and sources of support for the study reported?
Yes: Both competing interests and sources of support (financial or other) received for the study were reported; or the absence of any competing interest and source of support was acknowledged.
Partial: Either the competing interest or source of support was reported.
No: Neither competing interests nor sources of support were reported.

Note: Assessor(s) may decide to remove from the checklist the criteria that are not applicable to their project.