PRACTICAL APPLICATION

Challenges of Systematic Reviews of Economic Evaluations: A Review of Recent Reviews and an Obesity Case Study

Elisabet Jacobsen[1] · Dwayne Boyers[1] · Alison Avenell[2]

Abstract

Decision makers increasingly require cost-effectiveness evidence to inform resource allocation and the need for systematic reviews of economic evaluations (SREEs) has grown accordingly. The objective of this article is to describe current practice and identify unique challenges in conducting and reporting SREEs. Current guideline documents for SREEs were consulted and summarised. A rapid review of English-language SREEs, using MEDLINE and EMBASE, published in 2017/2018, containing at least 20 studies was undertaken to describe current practice. Information on data extraction methods, quality assessment (QA) tools and reporting methods were narratively summarised. Lessons learned from a recently conducted SREE of weight loss interventions for severely obese adults were discussed. Sixty-three publications were included in the rapid review. Substantial heterogeneity in review methods, reporting standards and QA approaches was evident. Our recently conducted SREE on weight loss interventions identified scope to improve process efficiency, opportunity for more transparent and succinct reporting, and potential to improve consistency of QA. Practical solutions may include (1) using pre-piloted data extraction forms linked explicitly to results tables; (2) consistently reporting on key assumptions and sensitivity analyses that drive results; and (3) using checklists that include topic-specific items where relevant and allow reviewers to distinguish between reporting, justification and QA. The lack of a mutually agreed, standardised set of best practice guidelines has led to substantial heterogeneity in the conduct and reporting of SREEs. Future work is required to standardise the approach to conducting SREEs so that they can generate efficient, timely and relevant evidence to support decision-making. We suggest only data extracting information that will be reported, focusing discussion around the key drivers of cost-effectiveness, and improving consistency in QA by distinguishing between what is reported, justified by authors and deemed appropriate by the reviewer.

1 Introduction

Economic evaluations (EEs) are an increasingly common requirement for evidence-based resource allocation decisions [1–3]. Systematic reviews of clinical effectiveness, utilities and costs are often used to populate cost-effectiveness models. However, systematic reviews of EEs (SREEs) are rarely used to provide cost-effectiveness evidence directly to policy makers.

✉ Elisabet Jacobsen
  elisabet.jacobsen@abdn.ac.uk

1 Health Economics Research Unit, University of Aberdeen, Polwarth Building, Foresterhill, Aberdeen AB25 2ZD, UK

2 Health Services Research Unit, University of Aberdeen, Aberdeen, UK

Key Points for Decision Makers

The need for systematic reviews of economic evaluation evidence has grown recently and is expected to continue to grow into the future as more decision-making bodies explicitly consider the value for money of new healthcare interventions.

A standardised approach on how to best summarise cost-effectiveness evidence is lacking and a set of mutually agreed best practice recommendations is required to improve the efficiency, relevance and transparency of future reviews.

The rationale behind SREEs and clinical-effectiveness reviews are similar. Decision makers are faced with often unmanageable amounts of information, from multiple different sources, often making conflicting value arguments. SREEs can provide an efficient mechanism to synthesise, report and disseminate the available cost-effectiveness evidence to answer a specific value question. By identifying where sufficient research is available to answer a policy question, and by demonstrating where the key research gaps lie, SREEs can help to avoid research waste, allowing funders and researchers to deploy their efforts more efficiently, where the greatest marginal gains of investment can be realised.

Despite their obvious benefit, there are many challenges to conducting a meaningful SREE. Some challenges are also encountered in clinical-effectiveness systematic reviews (e.g. heterogeneity in populations, study designs, methodology and consistency of outcome reporting), but others are unique to SREEs, and may be a limiting factor in their more widespread adoption. There is an extra layer of complexity in synthesising economic evidence due to the substantial variability across, and often within, different country-specific healthcare systems in terms of the payment for and delivery of healthcare. Therefore, different modelling assumptions (care pathways, time horizons, discounting procedures), analysis perspectives (person payer, insurance provider, government payer) or evaluation frameworks (cost-utility, cost-effectiveness, cost-benefit or cost-comparison) are needed to satisfy local decision-making requirements. Generalisability of economic evidence is difficult due to fluctuation of costs (due to inflation, costing practices, changes in payment for healthcare) and variation in society's valuation of health outcomes across countries and over time [4]. The substantial heterogeneity means that quantitative evidence synthesis using meta-analysis is often not possible and so narrative summaries of the evidence are usually required. Anderson [5] discussed the value of conducting SREEs, and how useful they are for answering specific policy questions, and suggest that the conduct of a SREE should be explicitly justified by reviewers, with a clear description of its purpose. Sculpher et al. [6] and Welte et al. [7] suggested 26 and 14 factors, respectively, that need to be transferable to another setting, in order for the EE to be relevant for decision-making. Boulenger et al. [8] pointed out that in order to be fully transferable, the economic model used would have to be tested with the data required for the specific context. The aim of this paper is to complement the existing research. We focus on how to address the challenges of a SREE given that the decision has been made to conduct one.

Large SREEs are resource intensive to complete and can generate large volumes of information, with much scope for inefficiencies in the process. This paper describes the current state of practice, using a rapid review of recently published SREEs, and discusses some of the unique challenges of conducting a SREE using a case study of weight-loss interventions. We envisage that the combination of a narrative review of current SREE practice and insights gained through our own large SREE experience will prove useful for future reviewers of EE evidence.

2 Objectives

The objectives of this paper are to describe the current state of practice, identify and discuss the key challenges of conducting SREEs, and suggest methods for improvement based on reflection of our experiences conducting a large SREE for weight loss interventions for severely obese adults (body mass index [BMI] ≥ 35 kg/m^2) as part of the REBALANCE (REview of Behaviour And Lifestyle interventions for severe obesity: AN evidenCE synthesis) project [9].

3 Methods

3.1 The REBALANCE Case Study

The REBALANCE project included a SREE for weight management (weight-loss and weight maintenance) programmes (WMPs), including diet and lifestyle interventions, behaviour change interventions, drug interventions and bariatric surgery for adults aged ≥ 18 years with severe obesity (BMI ≥ 35 kg/m^2).

Full details of the methods, including a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram, and results are reported elsewhere [9]. Briefly, studies conducted alongside trials and decision models that compared costs and outcomes in a single framework were included. A bespoke predefined online data extraction form was used to collect all reported economic data. Studies were assessed against the Drummond and Jefferson [10] and Philips et al. [11] checklists for trial-based and decision modelling studies, respectively.

Forty-six studies were included and narratively synthesised by modality of intervention (bariatric surgery [$n = 27$], lifestyle WMPs [$n = 16$] and orlistat [$n = 3$]) and EE type (within trial and decision modelling studies). Surgery was always deemed cost-effective due to long-term weight loss, but results for non-surgical WMPs were uncertain. The validity of the findings was dependent on the methodological quality of the studies. Common quality concerns were (1) a minority of non-surgical WMPs conducted long-term decision modelling to capture the full impact of obesity-related disease; (2) cost and utility implications from surgery-related complications were rarely included; and (3) weight regain

assumptions were rarely described or adequately tested in sensitivity analysis (SA).

3.2 Rapid Review of Systematic Reviews of Economic Evaluations (SREEs)

A rapid review of large (≥ 20 studies assessed) SREEs published in 2017 and 2018, recorded in two databases (MEDLINE and EMBASE) was undertaken. As the aim of the rapid review was to describe the most recent practice for conducting large SREEs, only the most recently published studies were included. The purpose of the rapid review was to provide context of current practice and it was not intended to be an exhaustive assessment of the literature. The search strategy, provided in Appendix 1.1 in the Electronic Supplementary Material combined filters for EE studies with a title search for "systematic review". Full-text, freely accessible articles published in the English language were included. Systematic reviews of cost-only studies, methods applications, conference abstracts and letters were all excluded. Abstract and full-text screening was conducted by one health economist using a predefined data extraction form in Microsoft Excel® (Microsoft Corp., Redmond, WA, USA). Information was collected on the number of items stated to be data extracted, actual number of items reported, quality assessment (QA) details, if any, and any visualisation of the results (e.g. graphical representation). A narrative discussion of current practice is provided, supplemented with simple frequency reporting of key information. The SREEs identified were not otherwise quality assessed.

After removal of duplicates, the search strategy identified 1663 potentially relevant titles. After review of titles and abstracts, 189 studies remained. A further 113 were excluded either because the full text was not retrievable or the article no longer met the inclusion/exclusion criteria (leaving 76 studies). A further 13 reviews failed to report the number of studies in the abstract and were also excluded. Sixty-three full-text papers were retrieved, analysed and data extracted. The median number of studies included in each SREE was 34 (range 21–115). The PRISMA diagram is presented in Appendix 1.2 in the Electronic Supplementary Material.

3.3 Review of Current Guidelines

Current guidelines for the conduct of SREEs were narratively summarised to identify current best practice methods. Details of some of the most widely cited current guidelines (e.g. Cochrane guidelines) were obtained from a hand search of reference lists of the identified studies from our rapid review of reviews and the reference lists of other published SREEs. The guidelines focused on were those that provided guidance on the methods of data extraction and reporting.

3.4 Description and Summary of the Key Challenges of Completing a SREE

The results that follow describe the experiences of conducting the REBALANCE SREE and the challenges encountered and lessons learned, and make recommendations for future practice across three main areas of the review process: (1) data extraction: what and how much data should be extracted, and how do we extract data efficiently; (2) evidence synthesis: how can we synthesise cost-effectiveness evidence in a meaningful manner, that is informative for decision makers, when meta-analysis is not possible; and (3) QA: how we can better quality assess studies, in a manner that is more consistent with the research question/disease area for review, and are there ways in which the relevance, usability and transparency of commonly used QA and reporting checklists can be improved? The experiences of the REBALANCE study are supplemented with the findings from the rapid review of current practice where appropriate.

4 Results

4.1 Data Extraction

4.1.1 Current Guidelines

The Centre for Reviews and Dissemination (CRD) provides broad guidance for undertaking SREEs, including advice to use predefined and piloted data extraction forms [12]. The guidance suggests categories of data (e.g. study results) but provides limited further guidance for reviewers. The *Joanna Briggs Institute Reviewers' Manual 2014* includes a pre-specified list of items to extract [13]. Van Mastrigt et al. [14] recommend piloting a data extraction form for thoroughness. Wijnen et al. [15] recommend a list of 35 items for data extraction and illustrative dummy tables for study characteristics and study results; however, they emphasise that the extracted items should reflect the most important items, not just the items in the template provided. There is undoubtedly a trade-off between the efficiency gains of bespoke extraction and the cross-review consistency of a predefined shopping list of data extraction items that must be considered by reviewers. Overall, of the few guidelines mentioned here, current guidance documents provide a range of advice to reviewers, generating marked differences in approaches taken across SREEs.

4.1.2 Findings from the Rapid Review

Table 1 shows the results from the rapid review on data extraction methodology. Studies included in the rapid review are presented in Appendix 1.3 in the Electronic Supplementary Material.

Our findings are similar to those of Luhnen et al. [16, 17], who found that 32% and 44% of studies used a standardised data extraction form, respectively. It is concerning that many studies failed to provide any details on the data extraction method [16]. Where detailed information was provided, some studies failed to report on all the data extraction items. Such reviews lack transparency and may be inefficient, as reviewers are extracting more data than necessary.

4.1.3 Lessons Learned from REBALANCE

The majority of data items extracted from the REBALANCE review were not reported. The decision on what data to report was based on text constraints, despite the ability to provide data in online appendices, and striking a balance between reporting relevant informative data and information overload. On reflection, the process was highly inefficient. A helpful and more efficient approach would have been to use a predefined dummy table of results to design the data extraction form. The decision should be based on what is relevant to the research question and what is needed for reporting.

Another challenge from the REBALANCE review was that many studies reported multiple SAs. We extracted all SA results across all studies. However, a more useful and efficient approach may be to predefine the most important SA that might be expected to impact on the incremental cost-effectiveness ratio (ICER). For example, in obesity studies, this might include the rate of weight regain after intervention delivery.

Table 2 provides a suggested template that could be used for data extraction of results information, explicitly linked to result reporting. The suggested template is for cost-effectiveness results only. Study characteristics such as methodological details would also need to be data extracted.

To ensure relevance, the form should be piloted on a small number of studies to resolve any remaining issues and ensure data collection items are sensible, complete and relevant.

The REBALANCE study used an online data extraction form, with automated checks to help minimise extraction errors and flag disagreement amongst reviewers. Combining the positive experience of using online forms with the lessons learned, selection of relevant information for extraction could substantially improve the efficiency of the review process.

4.2 Reporting and Presentation of Cost-Effectiveness Results

4.2.1 Current Guidelines

Reporting guidance focuses on providing transparent cost-effectiveness information. This includes reporting study methods such as perspective, population, time horizon, effectiveness source, benefit measurement, study type, model type, SAs (both a list of parameters and range of ICERs), and base-case deterministic and probabilistic results [18].

4.2.2 Findings from the Rapid Review

Table 3 describes the methods used for reporting study results in the rapid review.

Encouragingly, the majority of studies followed PRISMA recommendations, a key example of how clear guidance can improve standardisation of reporting. Visual presentations of results can help deliver key messages. However, caution is required as there is a risk that the reader may incorrectly compare results across studies without considering the appropriate caveats regarding heterogeneity. Luhnen et al. [17] also found that most studies conducted a narrative synthesis of results and did not use a graphical representation.

4.2.3 Lessons Learned from REBALANCE

Base-case ICERs were reported alongside minimum and maximum ICERs from SA for each study. However, the reporting could have been more informative and transparent if a template such as that suggested in Table 2 was used to derive the proportion of studies that conducted each key SA and the associated impact on the base-case ICER. For example, the REBALANCE de novo model found that results were highly sensitive to weight regain assumptions. However, only a minority of studies in the SREE specified their regain assumptions and fewer still tested them in a SA [9]. It is essential that SREEs identify the most important assumptions, and extract concise information on the impact of these assumptions on results. This would help to describe the true

Table 1 Rapid review findings on data extraction methods

Data extraction method	Proportion of studies [n (%)]
Detailed specification of data items provided	31 (49)
Qualitative summary of method, including statement that a predefined or standardised data extraction form was used without detailed description of the items	24 (38)
No details	8 (13)

Table 2 Possible template for reporting cost-effectiveness results

Study ID	Base-case results							Key drivers of cost-effectiveness (SA)						
	Int. cost	Δ cost	Δ outcome	Important base-case assumptions[e]	Base-case ICER	Prob. ICER	P (C/E)	Analysis[a]	Reported (Y/N)	Base-case value[b]	High value[c]	Low value[c]	ICER high[d]	ICER low[d]
								SA1						
								SA2						
								SA3						
								SA4						
								SA5						

Δ difference, *ICER* incremental cost-effectiveness ratio, *Int.* intervention, *N* no, *P(C/E)* probability of cost-effectiveness, *Prob.* probabilistic, *SA* sensitivity analysis, *Y* yes

[a]Pre-specified rank ordering of most important drivers of cost-effectiveness (SA1 should be the analysis considered a priori to be most important)

[b]The base-case parameter assumption

[c]The high/low parameter value varied in the SA

[d]The ICER that resulted from setting the parameter varied in the SA to the high/low value

[e]This should describe the most important drivers of cost-effectiveness. Specifying the base-case assumptions for these parameters will allow assessment of how important assumptions drive cost-effectiveness results across studies

uncertainty in cost-effectiveness results. We did not quantitatively synthesise the ICERs obtained from the REBALANCE review because of the significant heterogeneity across studies in terms of methodology, healthcare systems and definitions of interventions/comparators. Synthesising ICERs would run the risk of misinterpretation of the results, e.g. by comparing all ICERs to the same country-specific cost-effectiveness threshold.

4.3 Quality Assessment/Reporting Checklists

4.3.1 Current Guidelines

Multiple reporting and QA checklists exist and continue to develop over time. Currently, Cochrane recommends the Drummond and Jefferson [10] and Evers et al. [19] checklists for EEs conducted alongside single effectiveness studies and the Philips et al. [11] checklist for decision-analytical models [20].

4.3.2 Findings from the Rapid Review

Table 4 describes the different checklists identified as part of the rapid review, including a description of its general use (decision models or EEs alongside single effectiveness studies) and whether it can be considered a reporting standard or true QA tool.

Similarly to Luhnen et al. [16, 17], we find that the most popular checklist was the Drummond and Jefferson [10] checklist. It is concerning that approximately one-third of SREEs have not conducted any formal QA. Luhnen et al. [16] raised similar concerns, finding that about half of the health technology assessment (HTA) reports and one-third of the rapid reports did not provide a formal QA [16]. It is also concerning that many SREEs claim to have conducted QA, but further reading identifies that in fact reporting standards such as the CHEERS (Consolidated Health Economic Evaluation Reporting Standards) statement were used instead [22]. Furthermore, it was clear from the rapid review that many studies inappropriately used the same QA tool for model-based and trial-based EEs. None of the available QA tools contain sufficiently detailed questions to judge the quality of both study types. Furthermore, none of the checklists in Table 4 allow the reviewer to explicitly distinguish between reporting checks, justifications provided for assumptions and QA based on the reviewer's expertise. Reporting standards and QA tools are often incorrectly used inter-changeably in the assessment of EE studies.

In addition, some checklists include items containing multiple component subquestions. For example, were the assumptions about long-term health effects reported and justified? Or were costs and QALYs discounted appropriately? Such ambiguity can generate inconsistency among

Table 3 Method of reporting study results identified in the rapid review

Method	Proportion of studies [n (%)][a]
PRISMA/flow diagram	59 (94)
Trend graph of EE publications over time	7 (11)
Matrix	1 (2)
Graphical representation of the costs included or ICERs depicted	6 (10)
Figure of a typical model structure identified	1 (2)
Links between the included published models	1 (2)
Bar chart presenting study characteristics/study results	13 (21)
Reported key indicator(s) of cost-effectiveness	10 (16)

EE economic evaluation, *ICER* incremental cost-effectiveness ratio, *PRISMA* Preferred Reporting Items for Systematic Reviews and Meta-Analyses

[a]Some studies used different methods to present the study results; therefore, the numbers do not add up to 100%

Table 4 Reporting standards and quality assessment checklists

Checklist	No. of items	DM	SES	Reported[d]	Justified[d]	Quality assessed by the reviewer[d]	No. of studies in rapid review [n (%)][a]
Philips et al. [11]	57	X		X	X		7 (11)
Drummond and Jefferson [10]	35		X	X	X		16 (25)
Evers et al. [19]	19		X	X		X	8 (13)
QHES [21]	16		X	X	X		8 (13)
CHEERS [22]	24	X	X	X			11 (17)
NICE[b] [23]	20	X	X			X	2 (3)
European Network of Health Economic Evaluation Databases [24]	42	X	X	X			1 (2)
Stuhldreher et al. [25]	22		X	X			1 (2)
SIGN[c] [26]	12		X	X	X		1 (2)
No QA tool appears to have been used in the study							18 (29)

CHEERS Consolidated Health Economic Evaluation Reporting Standards, *DM* decision model, *NICE* National Institute for health and Care Excellence, *QA* quality assessment, *QHES* Quality of Health Economic Studies, *SES* single effectiveness study, *SIGN* Scottish Intercollegiate Guidelines Network

[a]Some studies included multiple checklists and some did not conduct a QA; therefore, the numbers to not add up to 100%

[b]On the NICE website, this checklist is not recommended for assessment of study quality or reporting quality [23]

[c]The SIGN checklist is based on the Drummond and Jefferson [10] checklist

[d] The columns 'Reported', 'Justified' and 'Quality assessed by the reviewer' refer to whether the checklists specifically asked about whether the item was reported, justified and/or quality assessed by the reviewer

reviewers. Walker et al. [27] reviewed ten checklists and found that they performed poorly in terms of inter-rater reliability. Time spent reconciling differences in opinion amongst reviewers could be minimised if questions were clearer.

4.3.3 Lessons Learned from REBALANCE

The lack of clarity in QA checklists made it challenging to consistently quality assess a large number of studies, with potential for within- and between-reviewer discrepancies. On reflection, the QA tools [10, 11] used in the REBAL-ANCE study did not always reflect the actual quality of the study. Some studies scored highly, whereas in fact they were judged by the reviewers to be of poor quality. For example, weight regain assumptions were poorly reported and justified by many studies in the REBALANCE SREE, yet they often scored highly using the checklists. One suggestion might be to adapt QA tools to include topic-specific bolt-on items that are pertinent to assessing the quality of EEs in specific topic areas, such as obesity.

Luhnen et al. [16] identified some examples of the use of bolt-on items to standard checklists, including adaptions to the Drummond and Jefferson [10] checklist by using additional QA criteria, e.g. based on a policy brief from

the Campbell Collaboration on economic methods [28] or WHO guidance on EEs for immunisation programmes [29]. Checklist adaptability can improve a checklist when it is able to better reflect the quality of the studies. The usefulness of the QA results depends on the quality and applicability of the QA checklists themselves.

5 Discussion

SREEs are growing in popularity, but the methodology is still developing. Our rapid review of reviews identifies some concerns regarding the lack of detail on the data extraction method, the potential risk of misinterpretation of the visually represented results and the lack of use of a formal QA. Luhnen et al. [17] have identified many similar concerns in their recently completed review, conducted in parallel with our work. Our study compliments the review of Luhnen et al. [17] by describing the lessons learned from our own recently completed large SREE as part of the REBALANCE study. We identify some solutions to common problems and make some practical recommendations for future practice to streamline and improve the review process.

There is limited consensus around guidelines to standardise practice. Evidence from the rapid review, and reflection on the REBALANCE SREE experience indicate scope to improve review efficiency. One suggestion to improve consistency of reporting and identification of important results is to link the results data extracted to the results reported, focusing on the most important, predefined SAs. Instead of data extracting all potentially relevant data, we recommend data extracting only what will be used for the report. For example, for the results, we recommend extracting only the data demonstrated in Table 2, i.e. not extract all SA, but only focus on predefined key drivers of cost-effectiveness and report those. This leads to time efficiency, while at the same time directing the reader to the most important drivers of cost-effectiveness.

QA checklists appear to lack clarity in their questions, and there appears to be substantial incorrect use of reporting checklists to assess EE quality. A helpful amendment to current instruments may be to simply distinguish between (1) what is reported; (2) whether it was justified by the authors; and (3) whether it was deemed of appropriate quality as judged by the reviewers. In addition, it may be helpful to supplement a core set of QA questions (e.g. discounting, time horizon, etc.) with topic-specific bolt-on items. For example, in obesity SREEs this might include questions about weight regain assumptions, appropriateness and completeness of the modelled obesity related diseases, and incorporation of costs and disutility associated with bariatric surgery complications. These bolt-on items should be SREE specific, agreed in advance of data extraction and linked where possible to the important analyses specified for data extraction (Table 2). To improve consistency in QA, it is important that journals and the health economic community adopt clear consensus guidelines for the conduct of SREE. This should be a priority of future research.

The SREE should aim to provide information on whether an intervention is cost-effective, but should be transparent about the likely significant heterogeneity across the included studies. Reviews should therefore provide ranges of plausible ICERs, alongside appropriate caveats about their generalisability across settings. SREEs should also provide key information from the included studies, including how they differ, in order to make the SREE as transparent and informative as possible. We caution against converting ICERs to a common currency as this might lead to misinterpretation of the results due to heterogeneity across healthcare systems. It is important to acknowledge the differences across studies and to comment on what impacts their generalisability [6, 7]. The process of transparency can be helped using graphical representation of results, but the associated caveats around heterogeneity should be clearly flagged. If applicable, the review should identify a subset of studies that answer the research question for the review. For example, if it is a review commissioned by the National Institute for Health and Care Excellence (NICE), then the review should identify the whole set of evidence, but also narratively (or otherwise) describe the subset of studies that are directly applicable to the research question. A SREE could also be useful in identifying decision models that could be adapted to address a specific research question, as opposed to building the de novo model from scratch. Trial-based EEs might not be able to answer the research questions because they are typically of short duration and based on a single study.

Current process guidelines for SREEs may provide some helpful advice, but further detailed consensus based guidance is needed. Shemilt et al. [30] argued that the Cochrane SREE methodology varies and has inconsistent reporting of the results. To improve transparency, it is good practice to publish the systematic review protocol [14], as was done with the REBALANCE review [31]. Guidelines on the conduct of the search for EE is well-developed (see the *Cochrane Handbook for Systematic Reviews of Interventions* [20]). Van Mastrigt et al. [14] also provide a step-by-step guide for reviewers developing their search. However, reviewers are faced with many challenges unique to SREEs that are not currently covered by guidance documents, leading to inefficient and inconsistent reviews.

A number of other studies have attempted to address some challenges of conducting SREEs. Pignone et al. [32] suggested ways forward for the SREE-specific challenges such as critically appraising studies and presenting results. For example, they recommend some key features to be critically appraised (including model type, perspective, costs

included and others) and they developed a set of guidelines for improving SREE [32], suggesting that ICERs for each comparison should be reported along with the input parameters to better understand the results. Different methods have been suggested on how to present the result in a simplified manner such as the three-by-three dominance ranking matrix (DRM) tool [14, 33], harvest plot [34] and hierarchical method [16, 35]. Others have discussed the use of the QA tools. Ramos et al. [36] highlighted the complexity and impracticality of checklists and suggested a 5-dimension framework on good practice procedure in modelling. Gomersall et al. [33] suggested separating multi-component questions into single items to avoid ambiguity.

There is still no consensus in the health economics community on which guidelines to follow when conducting SREEs. Consistency in SREE methodology would be highly beneficial; however, this is not possible without researchers or journals coming together and agreeing on a standardised approach to SREE.

Whilst we are transparent with regards to our rapid review methodology, it comes with some limitations. Our rapid review was confined to publication years 2017–2018 and only SREEs that included 20 studies or more. The review should therefore be considered as a description of current practice, rather than an exhaustive systematic assessment. The number of items in the predefined data extraction list and the number of reported items are subject to some uncertainty because often when the studies listed items to data extract, items were summarised, not giving a true picture of the predefined listed items versus reported items.

6 Conclusion

SREEs might prove to be challenging to conduct, with some studies debating its usefulness [5], but a well-conducted SREE is a useful framework for identifying good-quality economic evidence, avoiding research waste and identifying research priorities to support healthcare decision-making. Development of a detailed set of consensus-based guidelines for reviewers of EE evidence is a key priority for future research. We suggest an efficient approach to data extraction be taken and only data extract what will be used for the report, only report the key SAs (drivers of cost-effectiveness) in the results table, and improve consistency in the QA by distinguishing between what was reported, justified and deemed of appropriate standards by the reviewer.

Author Contributions All authors have reviewed the manuscript and approved of the final version to be published. Elisabet Jacobsen reviewed the cost-effectiveness evidence in the REBALANCE systematic review, reviewed current practice on conducting systematic reviews of economic evaluations, reviewed the literature on current guidelines for conducting systematic reviews of economic evaluations and wrote the paper. Dwayne Boyers supervised the REBALANCE review of cost-effectiveness evidence, conducted the rapid review of reviews search on current practice of conducting systematic reviews of economic evaluations and wrote the paper. Alison Avenell contributed to the reviews and critically reviewed the paper.

Compliance with Ethical Standards

Conflict of interest Elisabet Jacobsen, Dwayne Boyers and Alison Avenell declare that they have no conflict of interest.

Ethical approval This study did not require any ethical approval.

References

1. CADTH. Guidelines for the economic evaluation of health technologies: Canada. 4th ed. 2017. https://www.cadth.ca/sites/defau lt/files/pdf/guidelines_for_the_economic_evaluation_of_healt h_technologies_canada_4th_ed.pdf. Accessed 26 Apr 2019.
2. Australian Government Department of Health. Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee, version 5.0. 2016. https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf. Accessed 26 Apr 2019.
3. National Institute for Health and Care Excellence (NICE). Guide to the processes of technology appraisal. 2018. https://www.nice.org.uk/Media/Default/About/what-we-do/NICE-guidance/NICE-technology-appraisals/technology-appraisal-processes-guide-apr-2018.pdf. Accessed 26 Apr 2019.

4. Jensen CE, Jensen MB, Riis A, Petersen KD. Systematic review of the cost-effectiveness of implementing guidelines on low back pain management in primary care: is transferability to other countries possible? BMJ Open. 2016;6(6):e011042-2016-011042.

5. Anderson R. Systematic reviews of economic evaluations: utility or futility? Health Econ. 2010;19(3):350–64.

6. Sculpher M, Pang F, Manca A, Drummond M, Golder S, Urdahl H, et al. Generalisability in economic evaluation studies in healthcare: a review and case studies. Health Technol Assess. 2004;8(49):1–192 **(iii-iv)**.

7. Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics. 2004;22(13):857–76.

8. Boulenger S, Nixon J, Drummond M, Ulmann P, Rice S, de Pouvourville G. Can economic evaluations be made more transferable? Eur J Health Econ. 2005;6(4):334–46.

9. Avenell A, Robertson C, Skea Z, Jacobsen E, Boyers D, Cooper D, et al. Bariatric surgery, lifestyle interventions and orlistat for severe obesity: the REBALANCE mixed-methods systematic review and economic evaluation. Health Technol Assess. 2018;22(68):1–246.

10. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. BMJ. 1996;313(7052):275–83.

11. Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, et al. Review of guidelines for good practice in decision-analytic modelling in health technology assessment. Health Technol Assess. 2004;8(36):1–158 **(iii-iv, ix-xi)**.

12. Centre for Reviews and Dissemination. CRD's guidance for undertaking reviews in healthcare. Layerthorpe: York Publishing Services Ltd; 2009.

13. Gomersall JS, Jadotte YT, Xue Y, Lockwood S, Riddle D, Preda A. The Joanna Briggs Institute reviewers' manual 2014. The systematic review of economic evaluation evidence. Adelaide: The Joanna Briggs Institute; 2014.

14. van Mastrigt GA, Hiligsmann M, Arts JJ, Broos PH, Kleijnen J, Evers SM, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1/3). Expert Rev Pharmacoecon Outcomes Res. 2016;16(6):689–704.

15. Wijnen B, Van Mastrigt G, Redekop W, Majoie H, De Kinderen R, Evers S. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). Expert Rev Pharmacoecon Outcomes Res. 2016;16(6):723–32.

16. Luhnen M, Prediger B, Neugebauer EA, Mathes T. Systematic reviews of economic evaluations in health technology assessment: a review of characteristics and applied methods. Int J Technol Assess Health Care. 2018;34(6):537–46.

17. Luhnen M, Prediger B, Neugebauer EA, Mathes T. Systematic reviews of health economic evaluations: a structured analysis of characteristics and methods applied. Res Synth Methods. 2019;10(2):195–206.

18. Sullivan SM, Wells G, Coyle D. What guidance are economists given on how to present economic evaluations for policymakers? A systematic review. Value Health. 2015;18(6):915–24.

19. Evers S, Goossens M, De Vet H, Van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. Int J Technol Assess Health Care. 2005;21(2):240–5.

20. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. The Cochrane Collaboration; 2011. https://www.training.cochrane.org/handbook. Accessed 22 Apr 2019.

21. Chiou C, Hay JW, Wallace JF, Bloom BS, Neumann PJ, Sullivan SD, et al. Development and validation of a grading system for the quality of cost-effectiveness studies. Med Care. 2003;41(1):32–44.

22. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Cost Eff Resource Alloc. 2013;11(1):6.

23. National Institute for Health and Care Excellence (NICE). Guide to the methods of technology appraisal. 2013. https://www.nice.org.uk/process/pmg9/chapter/foreword. Accessed 30 May 2019.

24. Nixon J, Rice S, Drummond M, Boulenger S, Ulmann P, de Pouvourville G. Guidelines for completing the EURONHEED transferability information checklists. Eur J Health Econ. 2009;10(2):157–65.

25. Stuhldreher N, Konnopka A, Wild B, Herzog W, Zipfel S, Löwe B, et al. Cost-of-illness studies and cost-effectiveness analyses in eating disorders: a systematic review. Int J Eat Disord. 2012;45(4):476–91.

26. Scottish Intercollegiate Guidelines Network. Critical appraisal: notes and checklists. 2015. https://www.sign.ac.uk/checklists-and-notes.html. Accessed 22 Apr 2019.

27. Walker DG, Wilson RF, Sharma R, et al., editors. Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. Rockville: Agency for Healthcare Research and Quality; 2012.

28. Shemilt I, Mugford M, Byford S, Drummond M, Eisenstein E, Knapp M, et al. The Campbell Collaboration economics methods policy brief. Norwich: Campbell & Cochrane Economics Methods Group; 2008.

29. World Health Organization. WHO guide for standardization of economic evaluations of immunization programmes. Geneva: World Health Organization; 2008.

30. Shemilt I, Mugford M, Drummond M, Eisenstein E, Mallender J, McDaid D, et al. Economics methods in Cochrane systematic reviews of health promotion and public health related interventions. BMC Med Res Methodol. 2006;6(1):55.

31. REBALANCE trial team. (REBALANCE) REview of Behaviour And Lifestyle interventions forsevere obesity: AN evidenCE synthesis. 2016. https://www.fundingawards.nihr.ac.uk/award/15/09/04. Accessed 13 Nov 2019.

32. Pignone M, Saha S, Hoerger T, Lohr KN, Teutsch S, Mandelblatt J. Challenges in systematic reviews of economic analyses. Ann Intern Med. 2005;142(12_Part_2):1073–9.

33. Gomersall JS, Jadotte YT, Xue Y, Lockwood S, Riddle D, Preda A. Conducting systematic reviews of economic evaluations. Int J Evid Based Healthc. 2015;13(3):170–8.

34. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, et al. The harvest plot: a method for synthesising evidence about the differential effects of interventions. BMC Med Res Methodol. 2008;8(1):8.

35. Nixon J, Khan KS, Kleijnen J. Summarising economic evaluations in systematic reviews: a new approach. BMJ. 2001;322(7302):1596–8.

36. Ramos MCP, Barton P, Jowett S, Sutton AJ. A systematic review of research guidelines in decision-analytic modeling. Value Health. 2015;18(4):512–29.