Taylor & Francis
Taylor & Francis Group

Expert Review of Pharmacoeconomics & Outcomes Research

ISSN: 1473-7167 (Print) 1744-8379 (Online) Journal homepage: http://www.tandfonline.com/loi/ierp20

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)

BFM Wijnen, GAPG Van Mastrigt, WK Redekop, HJM Majoie, RJA De Kinderen & SMAA Evers

To cite this article: BFM Wijnen, GAPG Van Mastrigt, WK Redekop, HJM Majoie, RJA De Kinderen & SMAA Evers (2016) 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 Review of Pharmacoeconomics & Outcomes Research, 16:6, 723-732, DOI: 10.1080/14737167.2016.1246961

To link to this article: http://dx.doi.org/10.1080/14737167.2016.1246961

EXPERT REVIEW OF PHARMACOECONOMICS & OUTCOMES RESEARCH, 2016
VOL. 16, NO. 6, 723–732
http://dx.doi.org/10.1080/14737167.2016.1246961

Taylor & Francis
Taylor & Francis Group

REVIEW

OPEN ACCESS

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)

BFM Wijnen[a,b], GAPG Van Mastrigt[a], WK Redekop[c], HJM Majoie[b,d,f,g], RJA De Kinderen[a] and SMAA Evers [a,e]

[a]Department of Health Services Research, CAPHRI School of Public Health and Primary Care, Maastricht University, Maastricht, The Netherlands; [b]Department of Research and Development, Epilepsy Centre Kempenhaeghe, Heeze, The Netherlands; [c]Department of Health Policy and Management, Institute for Medical Technology Assessment, Erasmus University Rotterdam, Rotterdam, The Netherlands; [d]Department of Neurology, Academic Centre for Epileptology, Epilepsy Centre Kempenhaeghe and Maastricht University Medical Centre, Maastricht, The Netherlands; [e]Department for Economic Evaluations, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, The Netherlands; [f]School of Mental Health and Neuroscience, Maastricht University Medical Center, Maastricht, The Netherlands; [g]School of Health Professions Education, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands

ABSTRACT

Introduction: This article is part of the series "*How to Prepare a Systematic Review (SR) of Economic Evaluations (EE) for Informing Evidence-based Healthcare Decisions*" in which a five-step-approach for conducting a SR of EE is proposed.

Areas covered: This paper explains the data extraction process, the risk of bias assessment and the transferability of EEs by means of a narrative review and expert opinion. SRs play a critical role in determining the comparative cost-effectiveness of healthcare interventions. It is important to determine the risk of bias and the transferability of an EE.

Expert commentary: Over the past decade, several criteria lists have been developed. This article aims to provide recommendations on these criteria lists based on the thoroughness of development, feasibility, overall quality, recommendations of leading organizations, and widespread use.

ARTICLE HISTORY
Received 26 April 2016
Accepted 7 October 2016

KEYWORDS
Systematic reviews; economic evaluations; clinical practice guidelines; risk of bias; quality appraisal; data-extraction

1. Introduction

In this article, which focuses on Step 3 of the overall framework (see Figure 1 [1]), attention will be paid to data extraction, risk of bias assessment, and transferability when preparing a systematic literature review of economic evaluations (SR-EEs). Moreover, the article is also perfectly readable as a stand-alone read.

In 2003, the Appraisal of Guidelines for Research and Evaluation (AGREE) collaboration issued an instrument for evaluating the process of developing clinical practice guidelines (CPGs) and the quality of reporting [2]. Updated in 2010 to AGREE II [3], this instrument consists of several domains, of which 'rigor of development' and 'applicability' are of importance to this article. These two domains reveal that recommendations need to consider (1) health benefits, side effects, and risks (domain 3, item 11) and (2) the potential resource implications of applying the recommendations or, in other words, 'what does it cost when a certain recommendation is implemented?' (domain 5, item 20) [4]. This means that CPGs need to consider not only the potential effectiveness of health-care interventions but also their cost–effectiveness, as well as their overall impact on budget.

When creating or updating CPGs, SRs play a critical role in determining the comparative cost–effectiveness of health-care interventions, with the goal of creating an efficient health-care system [5]. However, EEs are prone to several biases. Bias occurs when there is a difference between the true value (in the population) and the observed value (in the study) from any cause other than sampling variability [6]. A bias can be unintentional or intentional and can have either substantial or little impact on the results of an EE [7]. In order to make optimum policy decisions, it is important to determine the risk of bias in an EE [7]. For example, EEs may have a perspective that is too narrow or may fail to incorporate important costs. In addition, one should be aware of the opportunity costs from decisions based on poor-quality EEs (i.e. misleading study findings, lack of transparency, and clarity in reporting [8]). Hence, in the past decades, several criteria lists have been developed to assess the risk of bias in EEs and to evaluate the transferability of EE. These lists are important tools that help to interpret and compare individual studies. However, due to the number and availability of all these tools, it can be difficult to make a careful selection.

Full EEs can differ in a variety of aspects, and all aspects can affect the quality of the evaluation and consequently bias results. The term 'full EE' refers to the comparative analysis of alternative courses of action in terms of both costs (resource use) and consequences (outcomes and effects) [5]. Basically, there are two approaches to performing an EE study

CONTACT BFM Wijnen b.wijnen@maastrichtuniversity.nl Duboisdomein 30, 6229 GT Maastricht, P.O. Box 616, 6200 MD, Maastricht, The Netherlands

Supplemental data for this article can be accessed here.

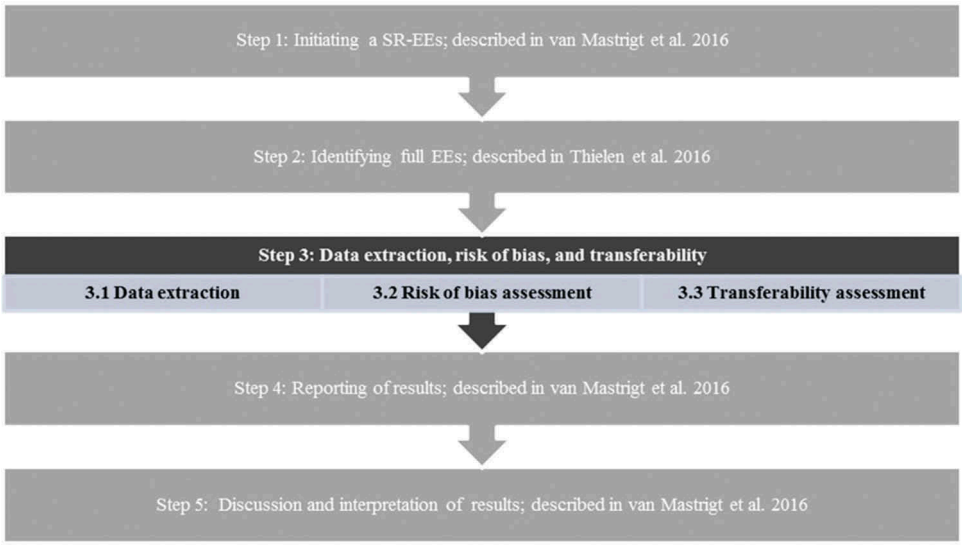

Figure 1. Overview of 5-step approach for preparing a systematic review of economic evaluations of healthcare interventions.

[7]: (1) an EE which is piggy-backed onto a clinical effectiveness study (e.g. a randomized controlled trial or observational study), often called a trial-based EE; and (2) a model-based EE, in which data from a wide range of sources (randomized controlled trials, observational studies, trial-based EEs, and other literature or reports) are combined using an economic model. Both are complementary to each other [5]. For a model-based EE, it is important that the external validation of the results, the key structural assumptions, and the data sources and derivation of the input data used in the model are well described. This way, potential policymakers or CPG developers are able to incorporate the strengths and limitations of the EE in their evaluation of the evidence [9]. It is important to keep in mind that the quality of EEs can be only as good as the quality of the trials on which they are based [9]. This holds true for both model-based EE and trial-based EE. As the field of effectiveness studies is relatively old in comparison with the field of EEs, methodological issues (i.e. the Cochrane collaboration's tool for assessing the risk of bias [10]), reporting standards (i.e. the consolidated standards of reporting trials (CONSORT) statement [11] and the strengthening the reporting of observational studies in epidemiology (STROBE) statement [12]), and the grading of evidence methods (i.e. the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) method for determining the quality of a body of evidence [13]) have been established and are regularly being used and recognized in the development of CPGs for SRs. For example, as in clinical studies, biases in (full) EEs can occur as a result of poor methodological quality, which can impact the validity of the results in terms of generalizability or transferability [7]. Evers et al. [7] have identified some methodological biases (so-called 'pretrial and during trial biases'), e.g. biases that occur as a result of a narrow perspective, inefficient comparator, cost measurement omission, or inappropriate discounting. In addition, they have identified some after-trial biases such as reporting and dissemination bias. Although these biases are thought of more in relation to trial-based EE, most of them are also applicable to model-based EE.

Therefore, this article will focus only on risk of bias and transferability checklists specifically tailored for the critical appraisal of EEs. SR-EEs can be categorized roughly into three groups: (1) multipurpose reviews, (2) reviews for informing the development of CPGs, and (3) reviews for developing decision analytic models. Both multipurpose SR-EEs and SR-EEs for guideline development aim to synthesize and critically appraise existing EEs of a health-care intervention or disease area in order to inform policy decisions [1,14]. The guidance in this article covers only the first two types of SR categories. Accordingly, this article is aimed mainly at CPG developers although the checklists can also aid others who want to prepare SR-EEs, like health technology assessment (HTA) researchers, systematic reviewers, and students, as they seek to identify the different steps, important key sources, and practical information to gain basic knowledge on this topic.

To be sure all relevant data of the included studies have been collected, it is important to develop a data extraction sheet for more systematic data collection. A data extraction sheet is an organized table in which all relevant items which need to be extracted for the review are listed; this needs to be completed for every study in order to collect data systematically. The inclusion of items depends on the research question or study objective and on the study design and outcomes predefined in the study protocol (see Step 1.3 of the overall framework in Van Mastrigt et al. [1]).

Accordingly, this article will first discuss the data extraction sheet and then present an overview of the methods most commonly used to assess the risk of bias and the transferability of EEs.

2. Step 3.1 of the overall framework: data extraction

This step entails extracting all relevant data from the included studies. For every SR-EE, a tailored data extraction sheet needs to be developed. Which items are included depends on the research question or study objective and on the study design and outcomes predefined in the study protocol (see Step 1:

'Initiating SR-EEs' of the overall framework). Consideration of the care pathway can be helpful in structuring the data extraction [15]. In addition, the risk of bias in the included studies needs to be appraised, in order to assess the possible impact of bias on the results of SRs (Step 3.2). Excel (Microsoft Office, Microsoft Corporation, Washinton) can be used for the digital registration of items. It is highly recommended that the extraction sheet be piloted for user-friendliness and completeness, using a few sample studies [15,16]. Then, if needed, the data extraction form can be adapted before starting data extraction of all studies. For Step 3.1, the data extraction of study characteristics, methods, and outcomes, it is important to simply report the findings as reported by the authors of the study and not draw any conclusions from them. This is in contrast with Step 3.2, the risk of bias assessment, in which a critical appraisal of the studies is necessary for answering all questions.

There are several example of data extraction forms available from the literature [15–17], containing many common items. These items can be classified in two groups. First, the general study characteristics: these are, for instance, author, year of publication, type of intervention, control treatment, eligibility criteria, study perspective, type of EE, and analytic approach (trial based versus model based). Second, the study methods and outcomes: these include resource use, costs, effects, measurement, valuation methods, incremental cost–effectiveness ratios, uncertainty analyses, sensitivity analysis, and conclusions. Based on our experience, we recommend including all relevant items from the list in Table 1 in the initial data extraction. If one is particularly interested in model-based EEs, one could extend this list with the external validation of the results, the key structural assumptions, and data sources and derivation of the input data used in the model. Using a picklist is recommended for choosing the different answers. Furthermore, in order to facilitate the interpretation of the results, a disaggregated presentation of the results, as well as incremental cost–effectiveness ratios, is highly recommended [15]. When presenting information derived from the data extraction, in some cases it may be more appealing to present in a table than others. Table 2 and Table 3 provide an example of how to present the general study characteristics and the economic evidence. In this study of De Kinderen et al. [18], a ketogenic diet is compared with care as usual to reduce epileptic seizures in children with intractable epilepsy.

3. Step 3.2 of the overall framework: risk of bias assessment

This step focuses on the risk of bias assessment for the studies included in SR-EEs. Although the risk of bias in EEs is equally important in CPG development and multipurpose reviews, differences might occur in the type of EEs included. In general, full EEs are recommended as being the most valid way to conduct an EE. Accordingly, we would like to stress that full EEs should be preferred over partial EEs at all times. However, in CPG development and/or in the absence of full EEs, one might be interested in partial EEs (e.g. costs analyses). Partial EEs may represent important intermediate stages in our understanding of the costs and consequences of health services programs and therefore might be convenient, e.g. in (early) CPG development [5]. Hence, both full and partial EEs will be discussed separately, with the difference that partial EEs will be discussed only in relation to CPG development.

In addition, although the risk of bias assessment and the way of reporting the results of EEs might seem like two distinct topics, in practice both topics are intertwined and difficult to differentiate from one another. For example, in order for a flawlessly conducted EE to be perceived as having 'low risk of bias,' it should be reported in a transparent and comprehensive way. While this article will focus on the risk of bias assessment of EEs, it is important to keep this in mind when reading the rest of the article. Specifically in order to assess the reporting quality of an EE, the International Society of Pharmacoeconomics and Outcomes Research (ISPOR) taskforce has developed the Consolidated Health Economic Evaluation Reporting Standards, in which recommendations are made to optimize the reporting of health EEs for all types of EE derived by either trial- or model-based EE [8].

3.1. Risk of bias assessment for multipurpose SR-EEs

As full EEs are considered to be the best strategy to answer efficiency questions [5], most checklists focus solely on full EEs. Over the past decades, several criteria checklists for the risk of bias assessment of full EEs have been constructed. A recent SR by Walker et al. [19] identified 10 checklists and criteria lists published between 1992 and 2011. In addition to the studies identified by Walker et al., we identified three additional studies: the checklists of Sculpher et al. [20], Philips et al. [21], and Caro et al. [22] (ISPOR checklist). An assessment of these checklists was made based on the purpose of the development, thoroughness of the development process, number of criteria checklists, operationalization of the questions, assessment instructions, time to complete, whether the checklist includes an overall quality score, and the number of references (providing us with an indication of its frequency of use). The full overview can be found in online supplementary material I. The British medical journal (BMJ) checklist [23] and the consensus on health economics criteria checklist (CHEC)-extended checklist, which is an extension of the original CHEC checklist to include a question regarding model-based EE [24–26], are commonly considered to have more scrutiny than most other lists [16]. Accordingly, the Cochrane collaboration recommends using one of these two checklists to assess the risk of bias of full trial-based EEs conducted alongside single-effectiveness studies. In addition, the BMJ checklist is also recommended by the Campbell & Cochrane Economics Methods Group for use in SRs. However, if the scope of the critical review of EEs encompasses relevant economic modeling studies, then assessments of the risk of bias of such studies will need to be informed by a different checklist, as the BMJ checklist and the CHEC-extended checklist are relevant but not sufficient for modeling studies [16]. Both the Cochrane collaboration and the National Institute for Health and Care Excellence (NICE) recommend using the Philips checklist to assess modeling studies [27]. However, as the Philips checklist contains a relatively large amount of criteria, using this checklist may not be feasible if one is interested specifically in a large number of model-based EEs. In cases where one is specifically interested in model-based EE and if the expected number of included studies is low (e.g. <10 studies; pragmatic decision), the Philips checklist [21]

Table 1. Items and explanation for the data extraction of economic evidence.

Number	Type of items	Explanation
	Checklist completed by	State the name of person who has filled out the data extraction sheet.
	General study characteristics	
1	First author and year of publication	Report: First author, title, journal name, publication date, volume, issue, page numbers, and link to the publication.
2	Sources of funding	Report: The source of funding cited in the article: write 'stated' or 'not stated' and specify. If any, give name(s) of organization or corporation. Specify if possible the source type (public research funds, NGO, government, academic/university, health-care industry, or other).
3	Competing interests	Competing interests: write "stated" or "not stated" and specify if any.
4	Publication type	Describe type of publication[b] (journal paper and HTA report)
5	Setting	List the country/countries, setting, and/or locations for economic evaluation.
6	Patient characteristics	Summarize inclusion and exclusion criteria (eligibility criteria/demographics).
7	Type of intervention	Describe the experimental treatment (service, program).
8	Control treatment	Describe the control treatment (service, program).
9	Eligibility criteria	Describe the eligible population and the population used for effect/cost data.
10	Study perspective	State the viewpoint of the analysis (society, health care, insurer, care provider, patient, and family).
11	Type of EEs	Specify the form of economic evaluation being used (e.g. CEA, CUA, CBA, and CCA).
12	Analytic approach	Describe the analytic approach: trial based or model based.
	Methods and outcomes of economic evaluations	
13	Time frame of the analysis (time horizon)	State the time horizon for both costs and benefits.
14	Discount rate	Was discounting performed?
15	Discount rate for costs	What was the discount rate for the costs(s)?
16	Discount rate for effects	What was the discount rate for the effects(s)? (i.e. the rate used to account for different timing of costs and effects)
17	Inflation rate	Was adjustment for inflation performed if unit costs stemmed from different years?
18	Reference year	What was the reference year of the analysis?
19	If model based	Detail any model used (Markov, Decision Tree, and Discrete Event Simulation)[c]
20	Type and category of costs	Describe the different cost types and categories used (e.g. direct in health care, indirect health care, and intangible costs).
21	Data source of resource use	Describe the data for resource use (e.g. clinical trials, administrative data, clinical databases, medical records, and published literature).
22	Methods for identifying resource use	Describe the methods used to identify resource use (questionnaire, survey, cost dairies, expert consultation, and formal consensus methods).
	Assumptions of the measurement of resources	Describe, for instance, method of imputation when incomplete measurement occurred.
22	Costs (in reported currency or in converted currency)	Present relevant costs and outcomes in both disaggregated and aggregated form (with confidence intervals and measures of significance).
24	Methods used to calculate unit costs	Describe the methods used to identify relevant unit costs (guidelines, own cost price calculations, and literature).
25	Costs[a]	Present relevant costs in disaggregated and aggregated form (with confidence intervals and measures of significance).
27	Data source of effects	Specify where utilities or benefits came from (literature values and elicited in the study).
28	Methods of measurement of effects	Specify source of effectiveness estimates (stated WTP, revealed WTP, and conjoint analysis).
29	Methods of valuation of effects	Specify methods of valuation of effects (indirect or direct measurement).
30	Effects[a]	Present relevant effects (utilities, [health] benefits, and outcomes) in disaggregated and aggregated form (with confidence intervals and measures of significance).
31	Incremental cost–effectiveness ratios	State the summary measure of benefit for CUA (e.g. QALY or DALY).
32	Analyses of uncertainty (e.g. sensitivity analyses)	Describe the analyses of uncertainty (e.g. statistical comparison, bootstrapping, sensitivity analysis [one-way, multiway], threshold analysis, analysis of extremes, and best/worst case analysis) and probabilistic sensitivity analysis.
34	Outcome(s) of analyses of sensitivity analyses	
35	Authors' conclusions	Report the conclusions of the authors

[a]When reporting the study outcomes, it is preferred to report the degree of uncertainty; therefore, in addition to reporting the mean (or median), a standard deviation (or range) should be reported [5].
[b]Duplicate publications of the same study need to be linked together.
[c]Check for details: Philips et al. [21].
CEA: cost–effectiveness analyses; CUA: cost–utility analyses; CCA: cost–consequence analyses; CBA: cost–benefit analyses; EEs: economic evaluations; QALY: quality-adjusted life year; NOG: Non-governmental organization; HTA: Health Technology Assessment; WTP: Willingness to pay; DALY: Disability adjusted life years.

Table 2. Example of data extraction for multipurpose review: general, RCT-related, and economic characteristics.

Author	Year	Disease	*N*	Intervention	Trial- or model-based EE	Comparator (s)	Outcome measure(s)	Perspective	Intervention cost-effective
De Kinderen et al.	2016	Refractory epilepsy	48	Ketogenic diet	Trial-based EE	Care as usual	QALY gain	Health-care perspective	No, based on cost per QALY
Etc.									
Etc.									

EE: economic evaluations; QALY: quality-adjusted life year; RCT: randomized controlled trial.
Based on De Kinderen et al. 2016.

Table 3. Example of data extraction for multipurpose review: cost–effectiveness results.

Author	Year	Disease	Main outcome	ICER/ICUR
De Kinderen et al.	2016	Refractory epilepsy	The results show that the KD reduces seizure frequency. The study did not find any improvements in quality of life and, therefore, unfavorable cost per QALY ratios resulted.	1) CEA: €18,422.21 per responder; 2) CUA: not presented but 5% probability of being cost-effective at the ceiling ratio of €50,000 per QALY
Etc.				
Etc.				

KD: ketogenic diet; CAU: care as usual; ICER: incremental cost–effectiveness ratio; ICUR: incremental cost–utility ratio; QALY: quality-adjusted life year; CEA: cost–effectiveness analyses; CUA: Cost-utility analysis; RCT: randomized controlled trial.
Based on De Kinderen et al. 2016.

could be used. In cases where the number of included model-based EEs is high (e.g. >10 studies; pragmatic decision), considering the feasibility and thoroughness of the developmental process, the ISPOR checklist is likely to be more practical for reviewing purposes [22].

In Table 4, an example is provided of how one can conduct the appraisal of a study using the CHEC-extended checklist. The appraisal process in this example was guided by assessment instructions specifically designed for the CHEC checklist. Again, the study of De Kinderen et al. [18] was used as an example. For most checklists, such instructions are available and make the appraisal process more straightforward.

3.2. Risk of bias assessment in SR-EEs for CPG development

The GRADE approach has been developed to rate the confidence in effect estimates (quality of evidence) for clinical outcomes and is often used and highly recommended in CPG development [13]. This approach was recently extended to include the quality of economic evidence (both for partial and full EEs). In general, the GRADE recommends that important differences in resource use should be included along with other important outcomes in the evidence profiles and summary of findings tables. In this process, four key steps have been identified: (1) identify items of resource use that may differ among alternative management strategies and that are

Table 4. Example of critical appraisal of the quality of the economic evaluation using the CHEC checklist.

	CHEC-extended items (Evers et al. 2005; Odnoletkova et al. 2014)	De Kinderen et al. (2015)
1	Is the study population clearly described?	Yes (see in method section, study population and sample size section, and result section; Table 2)
2	Are competing alternatives clearly described?	Yes, sufficient details provided on usual care group vs. ketogenic diet group.
3	Is a well-defined research question posed in answerable form?	Yes, last sentence of the introduction.
4	Is the economic study design appropriate to the stated objective?	Yes, an appropriate study design is used. The economic study design is a full economic evaluation (comparison of costs and effects of two interventions).
5	Are the structural assumptions and the validation methods of the model properly reported?	NA (trial-based EE)
6	Is the chosen time horizon appropriate in order to include relevant costs and consequences?	No, a 3-month time horizon is too short; for a societal perspective, a time frame of at least 1 year is generally accepted.
7	Is the actual perspective chosen appropriate?	Yes, the societal perspective is preferred in The Netherlands.
8	Are all important and relevant costs for each alternative identified?	Yes (see method section: protocol-driven intervention costs, healthcare costs, patient and family costs, and productivity costs are included.
9	Are all costs measured appropriately in physical units?	Yes, cost diaries were used.
10	Are costs valued appropriately?	Yes, sources of valuation including the reference year mentioned.
11	Are all important and relevant outcomes for each alternative identified?	Yes, outcomes are relevant, and they fit the research question and perspective.
12	Are all outcomes measured appropriately?	Yes, measurement instruments are relevant and described.
13	Are outcomes valued appropriately?	Yes (see method section) EQ-5D using a Dutch utility-tariff was used.
14	Is an appropriate incremental analysis of costs and outcomes of alternatives performed?	Yes, ICER/ICUR is calculated.
15	Are all future costs and outcomes discounted appropriately?	NA when using a time frame for analyses of less than 1 year, discounting is not needed.
16	Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis?	Suboptimal, not all parameters.
17	Do the conclusions follow from the data reported?	Yes, conclusions are supported by the data.
18	Does the study discuss the generalizability of the results to other settings and patient/client groups?	Yes, mentioned and (briefly) discussed in the discussion section of the article.
19	Does the article/report indicate that there is no potential conflict of interest of study researcher(s) and funder(s)?	Yes, the authors report having no conflict of interest.
20	Are ethical and distributional issues discussed appropriately?	Yes, the study was approved by the Medical Ethics Committee of Maastricht University.

Based on De Kinderen et al. 2016.
NA: not applicable; EE: economic evaluation; ICER: incremental cost–effectiveness ratio; ICUR: incremental cost–utility ratio; CHEC: consensus on health economics criteria checklist; EQ-5D: EuroQol 5-dimensions.

potentially important to patients and decision-makers; (2) find evidence for the differences in resource use between the options being compared; (3) rate the confidence in estimates of effect; and (4) if the evidence profile and summary of findings tables are being developed to inform recommendations in a specific setting, value the resource use in terms of costs for the specific setting for which recommendations are being made [28]. Resource use and the cost of treatments are included in the summary of findings tables. The cost–effectiveness estimates are included in the evidence profiles as background information. In this way, it is possible to include the results of partial EEs (e.g. cost analyses) in a systematic way in the GRADE approach when developing guidelines. However, the GRADE recommends excluding model-based EE as they are often based on trials which could lead to double counting [28]. Furthermore, the GRADE recommends that the confidence in effect estimates for each important or critical economic outcome should be appraised explicitly, using the same criteria as for health outcomes, so evidence derived from randomized trials starts at high quality, and evidence derived from observational studies starts at low quality [28]. In addition to integrating economic evidence in CPG development, using the previously described risk of bias checklist is complementary to the GRADE approach in assisting CPG developers in their deliberations [29]. Accordingly, to overcome the lack of compatibility with model-based EEs in the GRADE approach, the NICE has developed a checklist, specifically developed for the UK, composed of items from the CHEC and the Philips checklists. This composite checklist consists of 10 items regarding applicability and 12 items on study limitations (see Appendix H of the NICE Guidelines Manual [27]). One should be aware, however, that this list is based solely on expert opinion.

Overall, it can be concluded that when using the GRADE approach in developing CPGs, there is a way to systematically incorporate economic evidence from a partial EE. For further reading on this topic, we recommend checking the GRADE website (http://www.gradeworkinggroup.org/). In addition to incorporating economic evidence into CPG development, it is important to perform a complementary assessment on study applicability and the possible limitations for CPG using the NICE checklist [27]. As the NICE checklist is specifically designed for the UK, some minor adjustments are necessary to use it across jurisdictions, paying attention to such factors as the preferred perspective, the discount rate, or the preferred source of preference data.

4. Step 3.3 of the overall framework: transferability assessment of economic evaluations (for both multipurpose SR-EEs and SR-EEs for guideline development)

When conducting an SR to identify EEs which are applicable to a specific country or setting, or when developing a CPG and one is interested in cost–effectiveness or cost–utility data, it is important to determine the transferability and generalizability of such studies. Transferability is referred to as the extent to which the results of a study hold true for a different population or setting [30]. For example, results derived from a study conducted in a developed country will not be representative for use in a developing country. Generalizability is defined as the extent to which the results of a study can be generalized to the population from which the sample was drawn [30]. Although theoretically there is a clear difference between both concepts, they are often used interchangeably. To determine the transferability of a study, it is important to know what country-specific pharmacoeconomic guidelines exist and what the differences are between countries. To obtain information regarding country-specific pharmacoeconomic guidelines, the ISPOR has developed a comparative table of 33 countries, including key features for several (mostly European & American) countries (http://www.ispor.org/peguidelines/index.asp). For example, one should pay attention to the tariff used to derive quality-adjusted life years, which is one of the key transferability issues within cost–utility analyses, or to the perspective used and to the referred discount rates.

However, solely having the key features available for a country (or for a local setting) is not enough for most researchers to assess the transferability of an EE. Accordingly, as is the case for the risk of bias assessment, several instruments exist to evaluate the transferability of an EE (see online supplementary material II).

In an SR, Goeree et al. [31] identified seven checklists for determining the transferability of EEs. Based on the same criteria as the risk of bias checklists, an assessment of these checklists and the full overview can be found in online supplementary material II. In comparison to the risk of bias checklists, these checklists focus mainly on decision-making and on the implementation of study results in a particular setting. The checklist of Welte et al. was found to be a convenient list because it has clear cutoff points and can be used for the assessment of both trial- and model-based EEs. It has been applied successfully in the past [32], and the model has been tested extensively by Knies et al. [30]. The Welte checklist [33] consists of three general knockout criteria which need to be considered before proceeding to 14 specific knockout criteria (see Table 5). The Drummond (2009) checklist is largely based on the Welte checklist, and the two checklists differ only slightly in their application and content. Accordingly, overall, using the Welte checklist can be recommended. In addition, one should be aware that it is important to discuss the transferability of a particular study with clinicians as clinical practice might vary between countries.

In Table 5, an example is provided of how one may conduct the appraisal of a study using the Welte checklist. In this example, results of the study of De Kinderen et al. [18] are hypothetically transferred from The Netherlands to the UK setting. As can be seen from the example, the difference in perspective between the UK and The Netherlands may lead to the cost–effectiveness ratio being either too high or too low. In this case, one should recalculate the cost–effectiveness ratio excluding costs outside the health-care setting.

4.1. Usability of the different checklists for both multipurpose SR-EEs and SR-EEs for guideline development

We provide general recommendations regarding which checklist to use for assessing risk of bias and transferability.

Table 5. Example using the Welte checklist to determine the transferability of the study of the economic evaluation of the ketogenic diet from The Netherlands to the UK setting.

	Correspondence between study (The Netherlands) and decision country (UK)
General knockout criteria	
1. The evaluated technology is not comparable to the one that shall be used in the decision country.	NA
2. The comparator is not comparable to the one that is relevant to the decision country.	NA
3. The study does not possess an acceptable quality.	NA
	Correspondence between study (The Netherlands) and decision country (UK)
Methodological characteristics	
Perspective	Low (health care vs. societal)
Discount rate	Medium (3.5% UK vs. 4% NL)
Medical cost approach	High
Productivity cost approach	Low (friction costs method in NL vs. no productivity costs in UK
Health-care system characteristics	
Absolute and relative prices in health care	High
Practice variation	Medium
Technology availability	High
Population characteristics	
Disease incidence/prevalence	High
Case-mix	High
Life expectancy	High
Health-status preferences	High
Acceptance, compliance, and incentives to patients	High
Productivity and work-loss time	High
Disease spread	High

	Passed
	Passed
	Passed
	ICER of decision country based on ICER of study country is:
	Too low or too high
	Unbiased (short time horizon)
	Unbiased
	Too low or too high
	Unbiased
	Too low or too high
	Unbiased
	Unbiased
	Unbiased
	Unbiased
	Unbiased
	Unbiased
	Unbiased

NA: not applicable; ICER: Incremental cost–effectiveness ratio; NL: The Netherlands; UK: United Kingdom.
Based on De Kinderen et al. 2016.

These recommendations are based on a balance of the various aspects as described in the previous section. However, to determine what checklist fits best, several other study-specific characteristics determine the actual decision of which list or lists to select. These are, for instance, the time available for the review, the experience of the reviewers, the audience of the SR, and the purpose the checklist is designed for in relation to the aim of the review. In addition, the number of items and the time needed to complete a checklist are important factors in determining the applicability of a checklist. For example, the checklist of Philips [21] is often referred to by CRD as an instrument for appraising the risk of bias within modeling studies [15] although it is often ignored, due to the large number of criteria (61 items). Accordingly, it is important to look at feasibility when choosing the most appropriate checklist(s).

Moreover, one should be aware that raters are a relevant source of variability [34]; this highlights the importance of multiple raters (at least two) so that discrepancies can be resolved through consensus meetings between raters. In practice, this implies that, in addition to having multiple raters, a few studies (i.e. two or three) should be used to pilot the assessment between multiple raters, after which discrepancies should be discussed between raters to ensure a more uniform assessment strategy (see also Step 2.4, Thielen et al. [35] and Mastrigt et al. [1]).

5. Expert commentary

The starting point for the data extraction, risk of bias, and transferability assessment phase is the development of the data extraction sheet. This serves as a basis for collecting data from all the articles included under review. For convenience, one should include the selected risk of bias and transferability checklists in the data extraction sheet.

Next, as shown above, several checklists exist for assessing a variety of factors influencing the validity of study results within a particular setting. Accordingly, depending on the purpose of the review, different recommendations can be made. When one is interested in trial-based EEs, taking into account the thoroughness of the developmental process, the user friendliness, feasibility, and purpose of each checklist, the CHEC-extended [24] and the BMJ checklists [23] are most convenient to use. However, these checklists are insufficient when one is also interested in appraising model-based EEs. Therefore, although its length makes it cumbersome to apply to a large number of studies, the Philips checklist [21] should be considered. However, it should be noted that in current literature, there seems to be a lack of consensus regarding the best instrument for assessing the risk of bias of model-based EEs.

As stated above, all currently available checklists focus on full EEs, as full EEs are considered to deliver a high quality of evidence. However, especially in CPG, other factors might be considered in addition to cost–effectiveness data, such as the financial implications of the respective treatment. In this case, the use of a partial EE in an SR might be justified. The GRADE

approach came up with a method of incorporating economic evidence when developing CPG, but this method is not suitable for model-based EEs. In addition, the GRADE approach focuses on the estimated use of resources, which is only part of a (full) economic evaluation. Therefore, we would recommend performing a complementary assessment on study applicability and the limitations for CPG using the NICE checklist [27].

Looking at transferability, several checklists have been identified, of which the Welte checklist [33] has raised the most attention, likely due to the relative ease of application. In addition, the Welte checklist has been thoroughly examined [30], and the checklist of Drummond et al. [36] is based on the work of Welte et al. [33]. If one is particularly interested in assessing the applicability of HTAs to resource allocation decisions, the Grutters checklist [37] might be a suitable option (see online supplementary material II). One should keep in mind that, when incorporating economic evidence in a CPG, a transferability check should always be performed.

A summary of the recommendations made in this article can be found in Table 6.

The field of risk of bias assessment is developing quickly, resulting in numerous different checklists with different objectives. This article attempts to highlight the most important checklist currently available, but one should be aware of other checklists in this field. For example, although the product of their research is not defined as a checklist, Evers et al. [7] provide a list of risks of bias in trial-based EEs. Building on this and several other articles such as the Philips checklist [21], Adarkwah et al. [38] have developed the Bias in Economic Evaluation checklist (ECOBIAS), which is a checklist to determine the risk of bias in EEs. However, ECOBIAS is directed more toward model-based EEs. The checklist is aimed at providing a full overview of the biases that could occur in model- and trial-based EEs and includes a total of 22 biases, of which 11 are specific for model-based economic studies. Furthermore, for model developers or users of decision models, Vemer et al. [39] have developed a checklist, called 'Assessment of the Validation Status of Health-Economic decision models (AdViSHE),' which provides model users with a structured view into the validation status of the model, according to a consensus on what good model validation entails. AdViSHE may provide guidance towards additional validation of a model. However, when preparing an SR, using these checklists as add-ons to other risk of bias or transferability instruments will require a good understanding of EEs and will make the risk of bias assessment a time-consuming exercise.

6. Five-year view

Currently, data extraction in SR-EEs is done in several ways, and every author focuses on (slightly) different aspects. However, to improve the comparability of studies, there is a need for a more uniform standard with regard to data extraction sheets.

At this point, numerous checklists have been developed and applied within the field of SR and CPG development, specifically focused on EE. However, future research might

Table 6. Recommendations on data extraction, risk of bias, and transferability for a systematic review of EEs for SR-EEs for multipurpose and CPG development.

Step 3: Data extraction, risk of bias, and transferability	
Step 3.1: Data extraction for SR-EEs for multipurpose and CPG development	
Data extraction sheet	Adapt the data extraction for every specific study; include all relevant items from the list in Table 1.
	Include the selected risk of bias and transferability checklists in the data extraction sheet.
Data extraction	It is recommended to use a picklist to choose the different answers.
	Describe items to increase transparency of scoring.
	Disaggregated presentation of the results is highly recommended, as well as incremental cost–effectiveness ratios, in order to facilitate the interpretation of the results.
Step 3.2: Risk of bias assessment	
Risk of bias assessment for multipurpose SR-EEs and CPG development	Use CHEC-extended [24] or BMJ checklists [23].
	In cases where one is specifically interested in model-based EE and if the expected number of included studies is low (e.g. <10 studies; pragmatic decision), the Philips checklist [18] could be used.
	In cases in which the number of included model-based EEs is high (e.g. >10 studies; pragmatic decision), the ISPOR checklist is likely to be more practical for reviewing purposes [19].
Risk of bias assessment for CPG development	In general, full EEs should be preferred over partial EEs. However, in CPG development, partial EEs can be used to examine financial implications or the budget impact of a treatment.
	In the absence of a full EE, partial EEs may represent important intermediate stages in our understanding of the costs and consequences of health service programs and therefore might be convenient.
	Whenever possible, when developing CPGs, the GRADE approach should be used to systematically incorporate economic evidence of (partial) EE. In addition, it is important to perform a complementary assessment on study applicability and limitations using the NICE checklist [27].
Step 3.3: Transferability for SR-EEs for multipurpose and CPG development	
Transferability	Use the Welte checklist.
	It is important to discuss the transferability of a particular study with clinicians as clinical practice might vary between countries.
Usability for SR-EEs for multipurpose and CPG development	
Appraisal process	Raters are a relevant source of variability.
	It is important to have multiple raters (at least two) so that discrepancies can be resolved through consensus meetings.
	It is recommended that a few studies (i.e. two or three) should be used to pilot the assessment between multiple raters, after which discrepancies should be discussed to ensure a more uniform assessment strategy.

EEs: economic evaluations; SR: systematic review; CPG: clinical practice guideline; ISPOR: International Society of Pharmacoeconomics and Outcomes Research; GRADE: Grading of Recommendations Assessment, Development, and Evaluation; NICE: National Institute for Health and Care Excellence; BMJ: British medical journal; CHEC: Consensus on health economics criteria checklist.

further improve the risk of bias assessment of EEs. One important topic would be the development and validation of a single tool to assess and grade the risk of bias of both trial- and model-based full EEs in health care. Such a risk of bias assessment tool could make a substantial contribution to the field of health economics, as it would assist end users of cost–effectiveness studies to discriminate among the exploding body of literature and help producers of such studies to establish a clearer standard, potentially encouraging higher quality and greater rigor [19]. However, to achieve the necessary level of acceptance and use, the new tool must demonstrate validity and reliability [19]. Accordingly, it is expected that the GRADE approach will be adjusted to include model-based EE.

A third important topic would be more uniform and widespread guidance in the use of risk assessment instruments (e.g. which checklist should be used in what situation). An internationally supported protocol for the risk of bias assessment of EEs would support comparative analyses between reviews.

Fourth, there are only a limited number of studies looking at the reliability and validity of the discussed checklists. Future research should provide more insights into this matter.

The last topic would be to stress the need for increased transparency within the field of health economic model development, analysis, and reporting. This is particularly important for model-based EEs, where it is often difficult to fully grasp all important aspects of the model when reading only the accompanied article. By (freely) providing models (e.g. as online supplementary material), one could increase the transparency, and a more reliable risk of bias assessment could be conducted.

Key issues

- Currently, several checklists exist for assessing the risk of bias and the transferability of EEs.
- All these checklists focus on full EEs. However, when developing CPG, one might be forced to use a partial EE in resource allocation decisions. For this, the GRADE approach is highly recommended. In addition, it is important to perform a complementary assessment on study applicability and limitations using the NICE checklist [28].
- There is a lack of consensus regarding the best instrument for assessing the risk of bias within a model-based EE. Of the currently available checklists, the Philips checklist [21] is recommended when it is deemed feasible.
- The checklist of Welte et al. [33] should be used when determining transferability.
- There is a need to standardize the methods for data-extraction sheet development, use and filling in for EEs.
- There is a need for more uniform and widespread guidance in the use of risk of bias checklists (e.g. which checklist should be used in what situation). An internationally supported protocol for the risk of bias assessment of EEs would support comparative analyses between reviews.
- Future research should focus on the development and validation of a single tool to assess and grade the risk of bias in both trial- and model-based full EEs in health care.

Acknowledgments

The authors would like to thank the other members of the project: Laura Burgers (Erasmus University Rotterdam, Rotterdam), Jos Kleijnen (Maastricht University, Maastricht), Frederick Thielen (Erasmus University Rotterdam, Rotterdam), Toon Lamberts (Knowledge Institute of Medical Specialists, Utrecht), Wichor Bramer (Erasmus University Rotterdam, Rotterdam) for their valuable feedback on the draft of this paper. Furthermore, the authors would like to thank Barbara Greenberg for her English editing services and the master and bachelor students from the Maastricht University who were willing to provide us feedback on drafts of the paper.

Funding

This project was financially supported by the Dutch Society of Medical Specialists (Stichting Kwaliteitsgelden Medisch Specialisten): project number: 329666498, Utrecht, the Netherlands and the Netherlands Organisation for Health Research and Development (ZonMw), grant application number 836011018.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

ORCID

SMAA Evers http://orcid.org/0000-0003-1026-570X

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Mastrigt GAPGH, Arts JJC, Broos PH, et al. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: a five-step approach (part 1 of 3). doi:10.1080/14737167.2016.1246960. [Epub ahead of print].
2. Terrace L. Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care. 2003;12:18–23.
3. Brouwers MC, Kho ME, Browman GP, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. Can Med Assoc J. 2010;182(18):E839–E842.
4. cancer, C.p.a. AGREE II domains, items, examples and resources. [cited 2016 02 19]. Available from: http://www.cancerview.ca/cv/portal/Home/TreatmentAndSupport/TSProfessionals/ClinicalGuidelines/GRCMain/GRCAGREEII/GRCAGREEIIDomain3Rigour?_afrLoop=7202305453745000&jsessionid=24h456yAq7d254qi0lLI7VsmaXiA_PrK-x3k9VkKOzLOzM5cCIFs%21-1278458346&lang=en&_afrWindowMode=0&_adf.ctrl-state=y4y9grh4m_4
5. Drummond, M.F., Sculpher, M.J., Claxton, K, et al. Methods for the economic evaluation of health care programmes. Oxford: Oxford university press; 2015.
   • **Well-recommended handbook for anyone who is interested in economic evaluations.**
6. Brown GW. On small-sample estimation. Ann Math Statist. 1947;18(4):582–585.
7. Evers SM, Hiligsmann M, Adarkwah CC. Risk of bias in trial-based economic evaluations: identification of sources and bias-reducing strategies. Psychol Health. 2015;30(1):52–71.
8. Husereau D, Drummond M, Petrou S, et al. Consolidated health economic evaluation reporting standards (CHEERS) statement. BMC Med. 2013;11(1):80.

9. Rennie D, Luft HS. Pharmacoeconomic analyses: making them transparent, making them credible. JAMA. 2000;283(16):2158–2160.
10. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
11. Begg C, Cho M, Eastwood S, et al. Improving the quality of reporting of randomized controlled trials: the CONSORT statement. JAMA. 1996;276(8):637–639.
12. Von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Prev Med. 2007;45(4):247–251.
13. Schünemann H, Hill S, Guyatt G, et al. The GRADE approach and Bradford Hill's criteria for causation. J Epidemiol Community Health. 2011;65(5):392–395.
14. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J. 2009;26 (2):91–108.
15. Akers J, Aguiar-Ibáñez R, Baba-Akbari Sari A. CRD's guidance for undertaking reviews in health care. York (UK): Centre for Reviews and Dissemination (CRD); 2009.
    - **Well-recommended handbook for conducting a systematic review in general. Developed by the Center for Reviews and Dissemination, York University.**
16. Higgins JP, Green S. Cochrane handbook for systematic reviews of interventions. Vol. 5. Chichester: Wiley Online Library; 2008.
    - **Well-recommended handbook for conducting a systematic review in general. Developed by the Cochrane collaborations, an organization specialized in high-quality systematic reviews.**
17. Mlika-Cabanne N, Harbour R, De Beer H, et al. Sharing hard labour: developing a standard template for data summaries in guideline development. BMJ Qual Saf. 2011;20(2):141–145.
18. De Kinderen RJ, Lambrechts DAJE, Wijnen BFM, et al. An economic evaluation of the ketogenic diet versus care as usual in children and adolescents with intractable epilepsy: an interim analysis. Epilepsia. 2016;57(1):41–50.
19. Walker, D.G., Wilson, R.F, Sharma, R, et al. Best practices for conducting economic evaluations in health care: a systematic review of quality assessment tools. The Johns Hopkins University Evidence-based Practice Center. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012.
    - **Thorough review on quality assessment tools.**
20. Sculpher M, Fenwick E, Claxton K. Assessing quality in decision analytic cost-effectiveness models. Pharmacoeconomics. 2000;17 (5):461–477.
21. Philips, Z, Ginnelly L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment. Pharmacoeconomics. 2006;24(4):355–371.
22. Caro JJ, Eddy DM, Kan H, et al. Questionnaire to assess relevance and credibility of modeling studies for informing health care decision making: an ISPOR-AMCP-NPC good practice task force report. Value Health. 2014;17(2):174–182.
23. Drummond MF, Jefferson T. Guidelines for authors and peer reviewers of economic submissions to the BMJ. BMJ. 1996;313 (7052):275–283.
24. Evers S, Goossens M, De Vet H, et al. Criteria list for assessment of methodological quality of economic evaluations: consensus on health economic criteria. Int J Technol Assess Health Care. 2005;21(2):240–245.
25. Odnoletkova, I, Goderis, G, Pil, L, et al. Cost-effectiveness of therapeutic education to prevent the development and progression of type 2 diabetes: systematic review. J Diabetes Meta.b 2014;5:438 doi:10.4172/2155-6156.1000438
26. Consensus Health Economic Criteria - CHEC list Available from: https://hsr.mumc.maastrichtuniversity.nl/consensus-health-economic-criteria-chec-list
27. Excellence, N.i.f.C. Developing NICE guidelines: the manual; 2014; Available from: http://www.nice.org.uk/article/PMG20/chapter/7-Incorporating-economic-evaluation
28. Brunetti M, Shemilt I, Pregno S, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol. 2013;66(2):140–150.
29. Brunetti, M, Ruiz, F, Lord, J, et al. Grading economic evidence. Oxford: Wiley-Blackwell; 2010.
30. Knies S, Ament AJHA, Evers SMAA, et al. The transferability of economic evaluations: testing the model of Welte. Value Health. 2009;12(5):730–738.
31. Goeree R, He J, O'Reilly D, et al. Transferability of health technology assessments and economic evaluations: a systematic review of approaches for assessment and application. CEOR. 2011;3:89.
32. Essers BA, Seferina SC, Tjan-Heijnen VCG, et al. Transferability of model-based economic evaluations: the case of trastuzumab for the adjuvant treatment of HER2-positive early breast cancer in The Netherlands. Value Health. 2010;13(4):375–380.
33. Welte R, Feenstra T, Jager H, et al. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics. 2004;22(13):857–876.
34. Müller, D, Gerber-Grote, A, Stollenwerk, B, et al. Reporting health care decision models: a prospective reliability study of a multidimensional evaluation framework. Expert Rev Pharmacoecon Outcomes Res. 2016;16(5);1–9.
35. Thielen WF, Burgers LT, Bramer WM, et al., v.M.G.; How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: database selection and search strategy development. doi:10.1080/14737167.2016.1246962. [Epub ahead of print].
36. Drummond M, Barbieri M, Cook J, et al. Transferability of economic evaluations across jurisdictions: ISPOR good research practices task force report. Value Health. 2009;12(4):409–418.
37. Grutters JP, Seferina SC, Tjan-Heijnen VCG, et al. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. Value Health. 2011;14(5):777–784.
38. Adarkwah, C.C, van Gils, P.F, Hiligsmann, M, et al. Risk of bias in model-based economic evaluations: the ECOBIAS checklist. Expert Rev Pharmacoecon Outcomes Res. 2016;16(4);1–11.
39. Vemer, P, Ramos, I.C., van Voorn, G.A.K., et al. AdViSHE: a validation-assessment tool of health-economic models for decision makers and model users. Pharmacoeconomics. 2016;34:349. doi:10.1007/s40273-015-0327-2