BMJ Open

# Advocating uniformity in spine surgery: a practical disease-specific guideline for trial-based economic evaluations

Valérie N E Schuermans ,[1,2,3] Ruud Droeghaag ,[2,4] Sem M M Hermans,[2,4] Anouk Y J M Smeets,[1,3] Inge J M H Caelers,[1,2,3] Mickaël Hiligsmann ,[2,5] Wouter L W van Hemert,[4] Silvia Evers,[2,5] Henk van Santbrink,[1,2,3] The Expert Group

**To cite:** Schuermans VNE, Droeghaag R, Hermans SMM, *et al*. Advocating uniformity in spine surgery: a practical disease-specific guideline for trial-based economic evaluations. *BMJ Open* 2023;**13**:e073535. doi:10.1136/bmjopen-2023-073535

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2023-073535).

VNES and RD contributed equally.

VNES and RD are joint first authors.

Received 08 March 2023
Accepted 21 June 2023



For numbered affiliations see end of article.

**Correspondence to**
Dr Valérie N E Schuermans; valerie.schuermans@mumc.nl

## ABSTRACT

**Objectives** Despite the availability of general and national guidelines for the conduct and reporting of economic evaluations, there is heterogeneity in economic evolutions concerning spine surgery. This is partly the result of differing levels of adherence to the existing guidelines and the lack of disease-specific recommendations for economic evaluations. The extensive heterogeneity in study design, follow-up duration and outcome measurements limit the comparability of economic evaluations in spine surgery. This study has three objectives: (1) to create disease-specific recommendations for the design and conduct of trial-based economic evaluations in spine surgery, (2) to define recommendations for reporting economic evaluations in spine surgery as a complement to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 checklist and (3) to discuss methodological challenges and defining the need for future research.

**Design** A modified Delphi method according to the RAND/UCLA Appropriateness Method.

**Setting** A four-step process was followed to create and validate disease-specific statements and recommendations for the conduct and reporting of trial-based economic evaluations in spine surgery. Consensus was defined as >75% agreement.

**Participants** A total of 20 experts were included in the expert group. Validation of the final recommendations was obtained in a Delphi panel, which consisted of 40 researchers in the field who were not included in the expert group.

**Primary and secondary outcome measures** The primary outcome measure is a set of recommendations for the conduct and reporting, as a complement to the CHEERS 2022 checklist, of economic evaluations in spine surgery.

**Results** A total of 31 recommendations are made. The Delphi panel confirmed consensus on all of the recommendations in the proposed guideline.

**Conclusion** This study provides an accessible and practical guideline for the conduct of trial-based economic evaluations in spine surgery. This disease-specific guideline is a complement to existing guidelines, and should aid in reaching uniformity and comparability.

## STRENGTHS AND LIMITATIONS OF THE STUDY

⇒ This is the first available, practical guideline for disease-specific conduct of cost-effectiveness research in spine surgery.
⇒ The use of a modified Delphi method guarantees the support of professionals in this sector, which ensures a larger adherence and internalisation of these recommendations.
⇒ Although the expert group included international experts, the majority is from Europe, the guideline might thus reflect European preferences.
⇒ This guideline focuses solely on trial-based economic evaluations.

## INTRODUCTION

Taking into account ever-increasing healthcare expenses, the importance of economic evaluations is evident. Degenerative pathology is the main driver of costs within spine surgery.[1–4] The burden of degenerative pathology concomitantly increases with ageing of the population. To limit the increase of spine surgery-related healthcare costs, scarce healthcare resources should be allocated efficiently. Moreover, spine surgery has a direct influence on patient productivity, and an indirect effect on family and informal caregiver productivity. Hence, proper economic evaluation of surgical procedures is of utmost importance.[5 6]

The majority of recently published systematic reviews on economic evaluations in spine surgery conclude that there is abundant heterogeneity and a lack of quality within the field.[7–9] To investigate the extent of this heterogeneity, our group conducted a systematic review that assessed all trial-based economic evaluations in spine surgery as a first step of this Delphi process.[10 11] The aim was to evaluate the methodology and quality of all trial-based economic evaluations in spine surgery, which enabled us to identify the disparities in the current practice.[10 11]

The results of this broad systematic review show that the importance of economic evaluations is increasingly recognised, as reflected by the increase in the number of cost-effectiveness studies in the last decade.[11] The moderate quality and, more importantly,

extensive heterogeneity of these economic evaluations, however, greatly limit the comparability of their findings. Heterogeneity is caused by variable study designs, follow-up duration and outcome measurements such as utility, effectiveness and costs. Furthermore, studies differ largely in perspectives used, disparities in calculation methods of costs and/or charges, included cost items, different inclusion and exclusion criteria and baseline characteristics.[11] The results of this systematic literature review provide a foundation for the development of adequate recommendations to increase uniformity in economic evaluations in spine surgery.

Despite the availability of general and national guidelines for the conduct and reporting of economic evaluations, differing levels of adherence result in a wide variety of findings. A disease-specific guideline may provide more appropriate guidance in the conduct and reporting of economic evaluations in spine surgery.[12–14] General guidelines, by nature, do not incorporate disease-specific and topic-specific recommendations, which may provide insufficient guidance for specific topics. A disease-specific guideline as a supplement to general guidelines is necessary to ameliorate the overall quality and comparability of research.[8 15–17] Several disease-specific guidelines regarding the conduct of economic evaluations are available, but not in the field of spine surgery.[18–21]

Therefore, this study has three objectives: (1) to create disease-specific recommendations for the design and conduct of trial-based economic evaluations in spine surgery, (2) to construct recommendations for reporting economic evaluations in spine surgery as a complement to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 checklist[22] and (3) to discuss methodological challenges and defining the need for future research.

## METHODS

A modified Delphi study was conducted according to the RAND/UCLA Appropriateness Method.[23 24] A four-step process was followed to create and validate disease-specific statements and recommendations for the conduct and reporting of trial-based economic evaluations in spine surgery (figure 1). This study focuses on trial-based economic evaluations specifically. The majority of studies in the field of spine surgery are trial-based, and existing guidelines mainly focus on model-based economic evaluations. The aim is to create a practical guideline for clinical researchers in the field to help fill in the gap of application of trial-based economic evaluations. Authors formed a multicentre expert group consisting of experienced researchers in spine surgery and health economics.

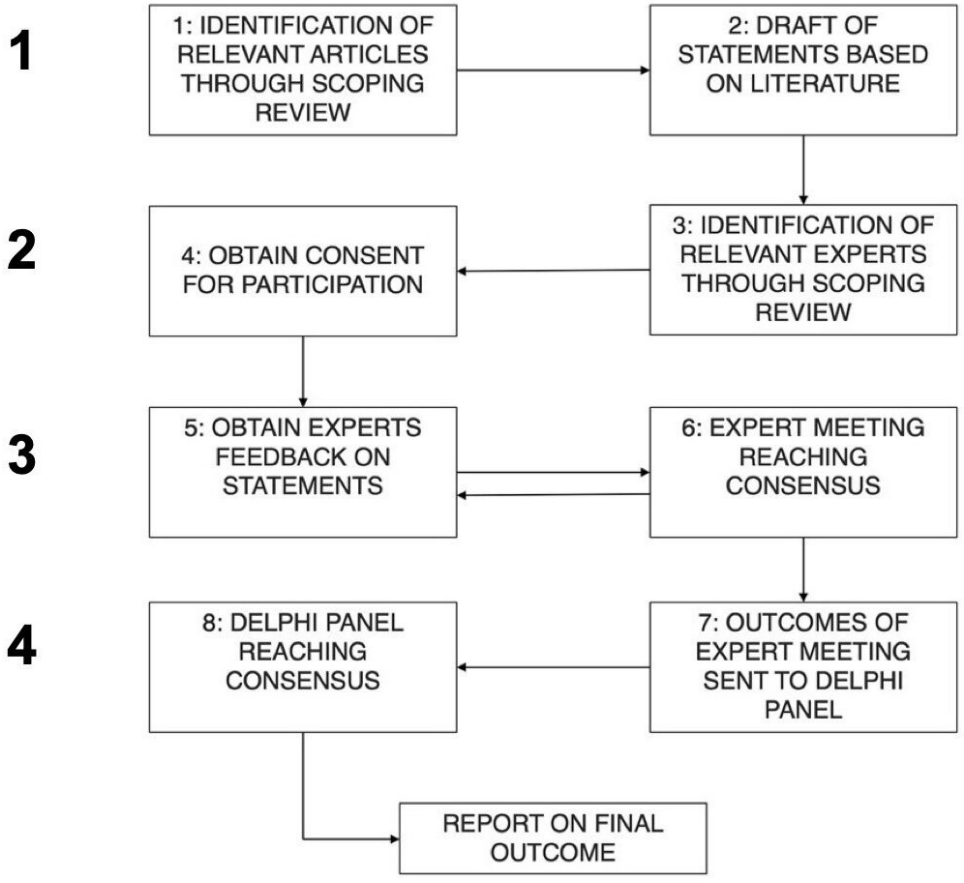


**Figure 1** Flowchart of the steps of the modified Delphi process.

## Systematic literature review and identification of experts

A systematic review was conducted in July 2021 to assess general guidelines or recommendations on economic evaluations, and articles concerning economic evaluations in spine surgery. The systematic review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement.[25 26] This will be made available as an Open Access article.[11]

### Identifying relevant studies

Relevant studies were selected and reviewed based on title and abstract. Articles deemed appropriate for inclusion were reviewed for further analysis. For more details and information, the full-text article can be consulted.[11]

### Identifying experts and Delphi panel

#### *Formation of the expert group*

Specifically first and last authors were identified from included articles derived from the systematic literature search to form the expert group. In addition, economic experts in health economics who contributed to the development of general guidelines and disease-specific guidelines were invited to join the expert group as well. To prevent missing relevant experts, the first and last identified authors were asked to propose additional suitable experts to be included. The role of the expert group was to perform a primary validation of statements drafted by the research group. All experts were approached for participation in the expert group through email. This email included a summary of the study design, the objectives and a request for participation. Written consent was obtained from all individual experts before participation. We aimed to include at least 15 experts. To ascertain an organised group discussion, we maintained a group maximum of 30 experts.

#### *Delphi panel formation*

To obtain a broader validation of the recommendations, a Delphi panel was formed with researchers in the field that were not included in the expert group. Whereas the expert group was formed based on the first and last authors of the articles included in the literature review, all identified authors of included articles could be included in the Delphi panel. Experts were also asked to propose additional colleagues, researchers and residents with experience in the field. The number of participants in the Delphi panel was not limited, a minimum number of 30 participants was required.

Expert group members were excluded from the Delphi panel. The Delphi panel was then asked to participate in an online survey

## Drafting first statements

The research group drafted statements based on the results of the abovementioned systematic literature review.[11] Recommendations were made for, but not limited to, the following topics:

1. Design and conduct of trial-based economic evaluations.
2. Reporting of economic evaluations, as a complement to the CHEERS checklist.
3. Discussion on methodological limitations and define the need for future research.

Full-text articles from the systematic review were analysed by the authors. The methodological features and limitations were extracted and collected in a spreadsheet divided into the abovementioned topics. All these features were synthesised into meaningful clusters and weighed by frequency and relevance. The first recommendations were drafted based on these findings. These drafts were then revised according to the feedback and input of the senior authors.

### Validation by expert group

#### *Online survey*

The previously developed statements were sent to the expert group to obtain a level of consensus and feedback. Feedback was received through a web-based questionnaire, built in Google Forms (online supplemental appendix 2). Demographic and professional characteristics of participants were collected. Level of consensus was assessed on a 0 to 10 scale for each statement, in which 0 meant 'disagree', 5 meant 'neutral' and 10 means 'agree'. The experts were asked whether they thought a statement was relevant to be included in the guideline on a scale from 0 to 3, in which 0 meant 'not relevant' and 3 meant 'relevant'. Experts were given the opportunity to provide textual feedback on each statement. Furthermore, all experts could suggest additional statements and were invited to leave further comments or advice for the research group. To prevent discussion between the experts, they were blinded during this stage of the process.

#### *Expert meetings*

Subsequently, two expert meetings were held to discuss statements and feedback provided by the expert group. The meetings were organised online with the use of Microsoft Teams [Version 1.5.00.27260]. The expert meetings were led by a member of the research group (VS or RD). Consensus was defined as a score of 75% or higher in terms of agreement in each category. A neutral score was not considered as disagreement. Statements were accepted if consensus was reached by the experts.[24 27] If consensus could not be reached on a proposed statement during the expert meeting, the statement was discarded, adjusted or reformulated. If no consensus could be reached after discussion, the statement was not included in the final guideline. After two expert meetings, consensus was reached on all drafted statements. These final statements were sent out to all participating experts for definitive approval.

## Validation by Delphi panel

The Delphi method is a structured process, commonly used to develop healthcare quality indicators and consists of four key components: iteration, controlled acquisition of

feedback, aggregation of responses and anonymity. We used the term modified as anonymity was not always applicable in our situation.[24 28] As described above, Google Forms was used (online supplemental appendix 3), recommendations were adjusted according to the feedback obtained in the expert meeting. For each recommendation, the Delphi panel could score 'Agree', 'Neutral', 'Disagree' or 'Don't know'.

All consensus statements were gathered and sent to the Delphi panel for final evaluation and validation. Statements reaching consensus of more than 75% were accepted for the final guideline after the two expert meetings and validation in the Delphi panel.

### Final report on outcomes

All consensus statements are reported in this paper, in the form of final recommendations for economic evaluations in spine surgery. Encountered methodological challenges and need for further research are discussed.

### Patients and public involvement

No patients were involved.

## RESULTS

### Drafting of statements

Forty-one statements were drafted by the research group based on the articles included in the systematic review of economic evaluations in spine surgery (n=108) and other relevant literature, including disease-specific or general guidelines (n=28). The initial statements can be found in online supplemental appendix 4. Feedback and input from the advisory board resulted in 35 statements remaining for expert group review.

### Expert group

Twenty-five experts who had extensive experience in spine surgery and/or cost-effectiveness research in the field of spine surgery agreed to participate in the expert group, of which 20 actively participated in either the online survey, the expert meeting or both. The group included experts from Europe (n=14), North America (n=4), Australia (n=1) and Asia (n=1). Seventeen of these experts had a doctorate, the remaining three had a University Master's degree. Eleven experts had a background in health economics, eight in medical science (spine surgeons) and one in biomedical engineering. The majority (n=14) of experts have been active in their field for over a decade, and a considerable number of experts have published over 50 articles in the last decade (n=8). All experts reported the use of general, national or regional guidelines in their current practice.

### Validation by expert group

During the first expert feedback round in the online survey, consensus was reached for 20 out of 33 statements (60.6%). Two additional statements (18 and 31) required multiple answers as they concern recommended effectiveness outcome measures to be used; level of agreement (LoA) could thus not be measured for these statements. The LoA after the online survey is summarised in figure 2. All statements were deemed relevant (figure 2). After thorough discussion during the two online expert meetings and adjustments of the statements according to the feedback, consensus was reached on 31 recommendations, including statements 18 and 31.

### Validation by Delphi panel

The 31 recommendations that reached consensus in the expert group were sent out to a larger Delphi panel for final validation. A total of 224 previously identified researchers in the field of spine surgery and/or health economics were invited to validate the recommendations through an online survey. A total of 40 researchers completed the survey. Consensus was reached for all recommendations, none of the recommendations reached more than 25% disagreement. Complete results can be found in figure 3.

### Final recommendations

A comprehensive overview of the final recommendations is provided in table 1.

The main elements of debate in the expert meetings are summarised per statement (#).

(#1) Although generally a randomised controlled trial is recommended, several experts pointed out that

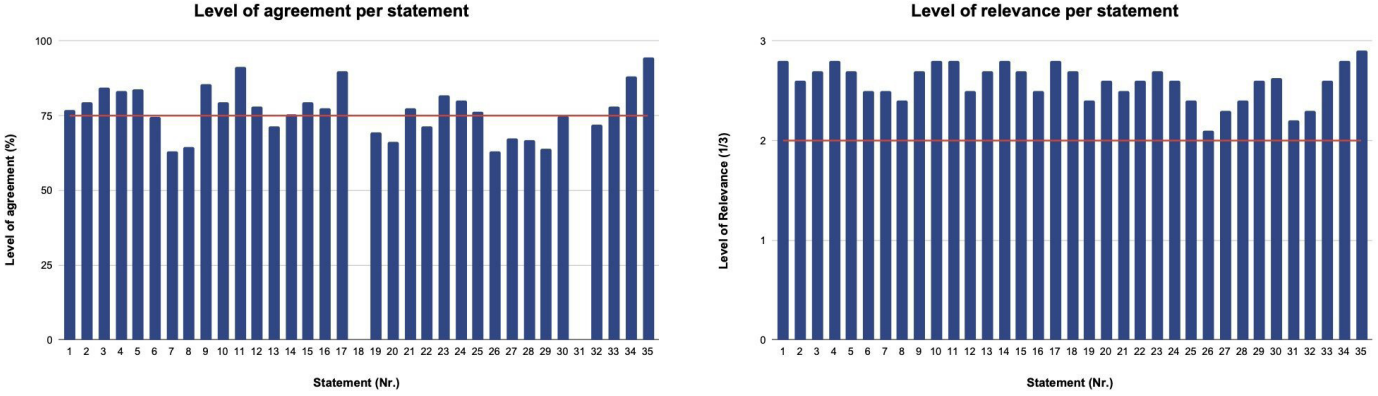


**Figure 2** Level of agreement per statement. Percentage (%) of agreement (left). Level of relevance (right) per statement, indicated with a score from 0 (irrelevant) to 3 (extremely relevant). The red line indicates the cut-off for consensus (>75% agreement).

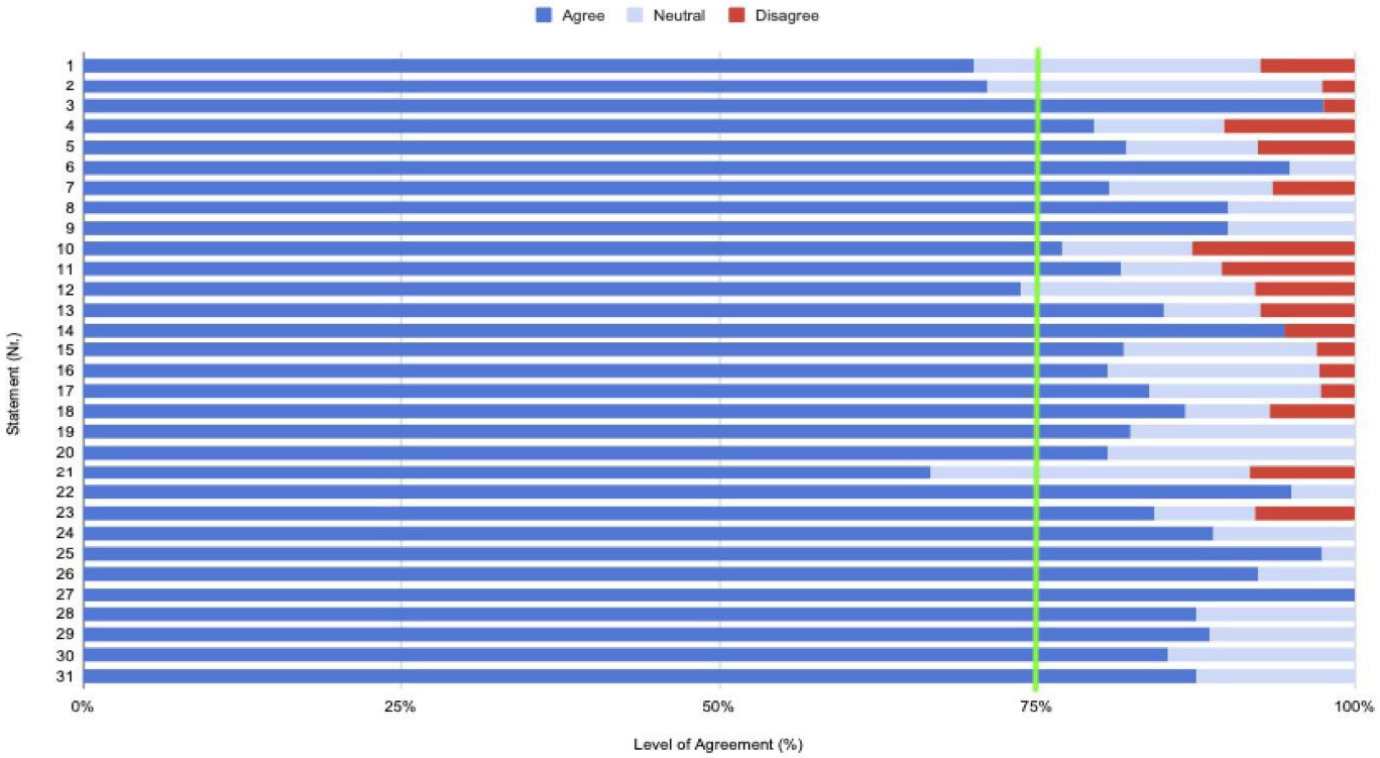


**Figure 3** Level of agreement on final statements. Indicated per statement in percentage (%). The green line indicates the cut-off for consensus (>75% agreement).

in specific cases, prospective comparative observational studies with very large cohorts are preferable.

(#2) As quality of life is the most important outcome for most spine surgeries, a cost-utility analysis (CUA) is preferred over a cost-effectiveness analysis (CEA). Since CEA investigates a specific clinical outcome of effectiveness, it is often too narrow to capture all relevant outcomes in a comparable fashion. However, in some situations in which a specific clinical outcome is of primary interest, a CEA is an acceptable alternative. We recommend the use of effect measures alongside utility measures in a CUA.

Choosing one preferred utility or effect measure is challenging. Researchers may prefer a specific outcome measure without solid scientific evidence. Choosing one effect or utility measure makes future studies more comparable. The chosen outcome measure in this guideline is the result of an extensive process under experts in the field. Hopefully resulting in consensus among future users.

(#5) Controversy remains concerning the definition of the standard of care. For example, the standard of care might differ per population, per country and over time. Therefore, it is important that the authors describe clearly how the standard of care is defined in the study.

(#7) Since discount rates vary per country or region, it was deemed better not to recommend a specific discount rate. Rather, it is recommended to consult national guidelines for discount rates. An additional analysis using a 0% discount rate is recommended to increase comparability between studies. As performing a sensitivity analysis was not within the scope of this work, it is advised to consult the Professional Society for Health Economics and Outcomes Research for further reading.[29]

(#11) We define complementary therapies in the clinical management pathway as all complementary therapies received both inside and outside of the hospital, as prescribed by the attending physician; for example, physical or occupational therapy. Costs of these therapies should be included in healthcare costs. All other self-initiated complementary therapies should be accounted for in the community costs when adapting a societal perspective, for example, physical or occupational therapy, acupuncture, etc.

(#17) Discussion persists regarding the optimal recall period for patient-reported outcome measures.[30] For accuracy, a short interval is preferable. However, for feasibility, longer recall periods are desirable. To optimise accuracy while maintaining feasibility, we recommend a recall period of 3 months. It should be noted that a recall period of 3 months does not necessarily mean that the questionnaire interval should also be 3 months.

(#21) Different questionnaires exist to evaluate quality-adjusted life years (QALYs). Several of these questionnaires can be suitable and are used in spine surgery research. As many of the existing studies used the EuroQol-5 Dimension (EQ-5D), and as the majority of the experts had a preference for the EQ-5D, we recommend this questionnaire to evaluate QALYs in a uniform fashion.

There was little discussion concerning the recommendations for reporting economic evaluations in spine

Table 1 Recommendations for trial-based economic evaluations in spine surgery

| Category | | Recommendations |
|---|---|---|
| General | | These recommendations are designed to supplement the existing (inter)national guidelines. If available, these should be consulted. All recommendations are designed for trial-based economic evaluations in spine surgery specifically. |
| Conduct | Study design | 1. Randomised controlled trials or meta-analyses of RCTs are the gold standard for measuring the effect size in economic evaluations. If an RCT is not feasible, a prospective study is preferred over a retrospective study, both with a comparative group. |
| | | 2. In spine surgery, cost-utility analysis is the preferred method. |
| | | 3. The economic evaluation should preferably be performed from both the healthcare and societal perspectives. |
| | | 4. In spine surgery, a minimum follow-up of 2 years is advised for clinical trials. A shorter follow-up period may be acceptable for a specific intervention, only if all costs and effects are expected within the chosen period. |
| | | 5. The standard of care should at least be chosen as comparator. If conservative treatment is the standard of care, this should be chosen as a comparator. |
| | | 6. An adequate time horizon should be adapted based on the interventions investigated and should be able to capture most of the relevant costs and benefits over time. |
| | | 7. Costs and effects should be discounted if a time horizon longer than 1 year is used. Sensitivity analyses for different discount rates should be performed, including an analysis with 0% discount rate. |
| | Outcomes (costs) | 8. Resources should be identified, measured and valued in detail, to ensure that the study can be replicated. |
| | | 9. Costs should be further divided into specified categories, that are more descriptive than direct and indirect costs. For example, healthcare costs, community costs, lost productivity, etc. |
| | | 10. In spine surgery, the following categories of costs should be included when adapting a societal perspective; healthcare costs, community costs, lost productivity, patient and family costs. When adapting a healthcare perspective, only healthcare costs should be included. |
| | | 11. All therapies of the clinical management pathway should be included in healthcare costs. If a societal perspective is adapted, all complementary therapies should be accounted for in the community costs. |
| | | 12. Resource use and medical consumption should be measured using existing databases of prospectively collected data. If such databases are not available or not, all relevant resources are covered, patient-reported measures can be integrated. Per patient resource data are preferred over the use of accumulated group data (eg, insurance data). |
| | | 13. Actual costs should be used. If costs are not available, tariffs (market prices) should be used. |
| | | 14. For the valuation of costs, of market prices, national guidelines or list prices and administrative data regarding hospital costs are recommended. |
| | | 15. For loss of productivity, both the friction approach and the human capital approach can be used. The chosen approach should be reported and justified. |
| | | 16. Mean national wages are preferred over self-reported wages. Self-reported wages could be used if the investigated population differs from the general population in terms of socioeconomic status. |
| | | 17. A maximum 3-month recall period for questionnaires and patient-reported outcomes regarding loss of productivity and resources used is advised. Other recall periods should be justified. |
| | | 18. A ‘steady state’ of the intervention should be assumed, costs should be estimated for routine employment. If relevant and applicable, costs and effects of learning and development could be included and should be reported separately. |
| | | 19. If national guidelines are not available, the used discount rate for costs and effects should be justified. |
| | Outcomes (utility and effect) | 20. Change in quality-adjusted life years (QALYs) is the most relevant outcome measure for economic evaluations in spine surgery. |
| | | 21. The EQ-5D-5L is the preferred patient-reported outcome questionnaire to determine utility outcome (QALY) in spine surgery. |

Continued

Table 1 Continued

| Category | Recommendations |
|---|---|
| | 22. Even though QALYs are of primary interest in cost-utility analysis in spine surgery, efficacy and safety outcomes (eg, pain, disability, adverse events) are relevant in most cases and should be assessed (table 2). |
| | 23. Lost productivity and informal (unpaid) care should be measured using existing databases of prospectively collected data. If such databases are not available or not all relevant resources are covered, patient-reported measures can be integrated. |
| Reporting | 24. If subpopulations are identifiable and relevant, post hoc analyses should be performed and reported. |
| | 25. The used categories of costs should be clearly reported. Costs should be reported separately per category. |
| | 26. All relevant efficacy and safety outcomes should be reported. |
| | 27. The reference year used for discounting of costs should be reported. |
| | 28. If differential discounting is used, the used rates and outcomes should be reported. |
| | 29. An incremental cost-effectiveness ratio (ICER) should be calculated and reported in all comparative studies. |
| | 30. A cost-effectiveness plane can be used to visualise cost-effectiveness. |
| | 31. A cost-effectiveness acceptability curve (CEAC) could be used to visualise the impact of willingness to pay for a certain outcome. |

RCT, randomised controlled trial.

surgery. Consensus was reached easily for all statements. As our recommendations on reporting in our guideline are complementary to the CHEERS checklist, we highly recommend adhering to this checklist.[22]

### Recommended outcome measures in CEAs

Throughout the expert meetings, experts were asked to suggest clinical outcome measures to be used in CEAs for different spinal pathologies. We categorised these into six domains: general, cervical spine, thoracic spine, lumbar spine, oncology and deformative pathology. Based on the experts' feedback, we defined a category of highly recommended outcome measures, defined as recommended by more than 50% of the experts. Optional outcomes measures consist of the remaining proposed outcome measures that can be considered when they are of specific interest (table 2).

## DISCUSSION

The objective of this international Delphi study was to create evidence-based recommendations to provide guidance to those involved in research trial-based economic evaluations in spine surgery. We successfully engaged a wide community of experts in the field to ensure that the final recommendations reflect participants' opinions, are meaningful, and help bridge existing gaps in practice. This has resulted in a set of 31 recommendations for the design, conduct and reporting of trial-based economic evaluations, as a complement to the existing guidelines. Moreover, we have identified and discussed methodological challenges and the need for future research.

Widespread variations in study possibly result from differing levels of adherence to the existing general guidelines. By defining these disease-specific guidelines, we aim to increase adherence and hence standardisation

Table 2 Recommended outcome measures for CEAs

| Domain | Highly recommended outcome measures | Optional outcome measures |
|---|---|---|
| General | Adverse events, reoperations, complications, VAS/NRS, ODI, COMI | Blood loss, OR time, LoS, HADS, MCS, PCS, Odom Criteria, GPE |
| Cervical spine | VAS neck/arm, mJOA | NDI |
| Thoracic spine | mJOA | EMS, Frankel Scale |
| Lumbar spine | VAS back/leg, RMDQ | – |
| Oncology | VAS axial pain, KPS, survival | OSRI, Bartels Score, ambulatory status |
| Deformative pathology | SRS | – |

CEAs, cost-effectiveness analyses; COMI, Core Outcome Measures Index; EMS, European Myelopathy Scale; GPE, Global Perceived Effect Score; HADS, Hospital Anxiety and Depression Scale; KPS, Karnofsky Performance Scale; LoS, length of stay; MCS, mental component summary; mJOA, modified Japanese Orthopedic Association; NDI, Neck Disability Index; NRS, Numeric Rating Scale; ODI, Oswestry Disability Index; OR, operation room; OSRI, Oswestry Spine Risk Index; PCS, physical component summary; RMDQ, Roland-Morris Disability Questionnaire; SRS, Scoliosis Research Society; VAS, Visual Analogue Scale.

in this kind of research. Although partly overlapping with the existing general guidelines, these spine-specific recommendations complement the general guidelines in several ways. First of all, standardisation of spine-specific utility, effectivity and cost measures will enlarge the uniformity of the outcome measures in cost-effectiveness research. Additionally, the Delphi method guarantees the support of professionals within this sector, which ensures a larger adherence and internalisation of these recommendations.

This spine-specific guideline is more extensive than the general guidelines for cost-effectiveness research. Although several of the statements might seem self-evident to some researchers, we aimed to provide a benchmark for all researchers in the field. This spine surgery-specific guideline for economic evaluations gathers all necessary features, making it accessible and easy to use for clinical researchers. Another important aspect is the awareness of the existence of these guidelines. Through the Delphi approach, both health economic and medical experts are informed of the existence of a disease-specific guideline in this overlapping field. Publication and implementation of this guideline create an opportunity for unified practice for the benefit of our patients.

The final recommendations are designed to supplement the existing (inter)national guidelines, which should always be consulted. All recommendations are designed for trial-based economic evaluations in spine surgery specifically.

## Strengths and limitations

The most important feedback from the expert group discussions was used to modify the recommendations and is presented in the paper. However, this paper does not incorporate all considerations to reject or support recommendations. Moreover, we only obtained textual feedback from the expert group, but not from the Delphi panel, as the aim was merely to measure the level of agreement in this group. Similarly, not all experts could attend the same meeting due to time zone differences, which might have influenced the discussions. Although the expert group included international experts, the majority were from Europe.

Our findings help define the few areas of ongoing controversy that can now be investigated with further focused studies. It is debatable whether generic tools, like EQ-5D-5L or Short Form 36 (SF-36), are optimal for measuring spine-related QALYs. The Core Outcome Measures Index (COMI) for back was developed with the aim to assess main outcomes of importance for patients with spinal pathology. However, the COMI is not yet validated to quantify changes in QALY and some discussion exist concerning the lack of consideration of mental well-being. Development of a spine-specific QALY tool could give better insight into spine-related quality of life. Since the score is relatively new, we did not find this outcome measure in the existing literature, however, we believe this to be a good effect outcome measure to use. To maintain comparability with other pathologies, this should be used alongside generic tools. As this guideline focuses solely on trial-based economic evaluations, the next step would be to provide disease-specific recommendations for model-based economic evaluations in spine surgery. This could provide a standardised, disease-specific reference case and in-depth recommendations for sensitivity analyses. We intended to incorporate live voting to measure consensus at conferences. As a consequence of the COVID-19 pandemic, the majority of conferences were virtual or postponed. Therefore, we opted for an online survey for the Delphi validation.

## Conclusion

This Delphi consensus study provides an accessible and practical guideline for the conduct of trial-based economic evaluations in spine surgery as a complement to existing guidelines. The final guideline includes 31 recommendations on the conduct and reporting of these economic evaluations. This guideline can be used as a checklist that serves as a minimum standard and should aid in reaching uniformity and comparability.

**Author affiliations**
[1]Department of Neurosurgery, Zuyderland Medical Centre Heerlen, Heerlen, The Netherlands
[2]Caphri School of Public Health an Primary Care, Maastricht University, Maastricht, The Netherlands
[3]Department of Neurosurgery, Maastricht Universitair Medisch Centrum+, Maastricht, The Netherlands
[4]Department of Orthopedic Surgery and Traumatology, Zuyderland Medical Centre Heerlen, Heerlen, The Netherlands
[5]Department of Health Services Research, Maastricht University, Maastricht, The Netherlands

**Collaborators** We would like to express our sincere gratitude for the valuable input by the expert group: Anne Broekema, Axel Svedbom, Bart Depreitere, Carmen Dirksen, Carmen Selva Sevilla, Daniel Cher, Elske Van den Akker – Van Marle, Francesca Fiorentino, Henk van Santbrink, Jordana Schmier, Jos Kuijlen, Kevin Ong, Laura De La Torre Perrez, Lei Si, Marieke Krol, Mickaël Hiligsmann, Silvia Evers, Sukhvinder Kalsi-Ryan, Takahito Fujimori, Wouter van Hemert.

**Contributors** Each author has played a crucial role in the development of the project. RD, VS, WvH, SE and HvS have contributed to the conception of the study. RD, VS, MH, SE and HvS have been involved in the design phase. RD, VS, SH, AS and IC have played a key role in data acquisition. RD and VS have performed the analysis. The interpretation of the results has been a collaborative effort involving RD, VS, MH, WvH, SE and HvS. Substantial revisions have been made by RD, VS, SH, AS, IC, MH, WvH, SE and HvS. Lastly, the supervision of the project has been carried out by MH, WvH, SE and HvS. These contributions demonstrate the collaborative nature of our work and the shared responsibilities among the authors. HvS is the guarantor of the study.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

**Patient consent for publication** Not applicable.

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

**Data availability statement** Data are available upon reasonable request.

**ORCID iDs**
Valérie N E Schuermans http://orcid.org/0000-0002-1733-8564
Ruud Droeghaag http://orcid.org/0000-0003-0955-7983
Mickaël Hiligsmann http://orcid.org/0000-0003-4274-9258

## REFERENCES

1 United Nations - Population Division Department of Economic and Social Affairs. *World population ageing 2017*. 2017.
2 Saifi C, Cazzulino A, Laratta J, *et al*. Utilization and economic impact of Posterolateral fusion and posterior/Transforaminal lumbar Interbody fusion surgeries in the United States. *Global Spine J* 2019;9:185–90.
3 de Kunder SL, Rijkers K, Caelers IJMH, *et al*. Lumbar Interbody fusion: a historical overview and a future perspective. *Spine (Phila Pa 1976)* 2018;43:1161–8.
4 Saifi C, Fein AW, Cazzulino A, *et al*. Trends in resource utilization and rate of Cervical disc Arthroplasty and anterior Cervical Discectomy and fusion throughout the United States from 2006 to 2013. *Spine J* 2018;18:1022–9.
5 Quah E, Mishan EJ, Quah E. cost-benefit analysis. 2007.
6 Hopkins B, Mazmudar A, Kesavabhotla K, *et al*. Economic value in minimally invasive spine surgery. *Curr Rev Musculoskelet Med* 2019;12:300–4.
7 Schuermans V, Smeets A, Boselie AFM, *et al*. n.d. Cost-effectiveness of anterior surgical decompression surgery for Cervical degenerative disc disease: a systematic review of economic evaluations. *Eur Spine J*
8 Droeghaag R, Hermans SMM, Caelers I, *et al*. Cost-effectiveness of open Transforaminal lumbar Interbody fusion (OTLIF) versus minimally invasive Transforaminal lumbar Interbody fusion (MITLIF): a systematic review and meta-analysis. *Spine J* 2021;21:945–54.
9 Caelers IJMH, de Kunder SL, Rijkers K, *et al*. Comparison of (partial) economic evaluations of Transforaminal lumbar Interbody fusion (TLIF) versus posterior lumbar Interbody fusion (PLIF) in adults with lumbar Spondylolisthesis: A systematic review. *PLoS One* 2021;16:e0245963.
10 Droeghaag R, Schuermans VNE, Hermans SMM, *et al*. Evidence-based recommendations for economic evaluations in spine surgery: study protocol for a Delphi consensus. *BMJ Open* 2021;11:e052988.
11 Droeghaag R, Schuermans VNE, Hermans SMM, *et al*. Anarchy in economic evaluations in spine surgery: a systematic review and qualitative assessment [BMJ Open 2023]. *BMJ Open* 2023;13:e067871.
12 Husereau D, Drummond M, Petrou S, *et al*. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Eff Resour Alloc* 2013;11:6.
13 Caro JJ, Briggs AH, Siebert U, *et al*. Modeling good research practices - overview: A report of the ISPOR-SMDM modeling good research practices task Force-1. *Med Decis Making* 2012;32:667–77.
14 Sanders GD, Neumann PJ, Basu A, *et al*. Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: second panel on cost-effectiveness in health and medicine. *JAMA* 2016;316:1093–103.
15 Kepler CK, Wilkinson SM, Radcliff KE, *et al*. Cost-utility analysis in spine care: a systematic review. *Spine J* 2012;12:676–90.
16 Alvin MD, Miller JA, Lubelski D, *et al*. Variations in cost calculations in spine surgery cost-effectiveness research. *Neurosurg Focus* 2014;36:E1.
17 Radcliff K, Guyer RD. Economics of Cervical disc replacement. *Int J SPINE Surg* 2020;14:S67–72.
18 Hiligsmann M, Reginster J-Y, Tosteson ANA, *et al*. Recommendations for the conduct of economic evaluations in osteoporosis: outcomes of an experts' consensus meeting organized by the European society for clinical and economic aspects of osteoporosis, osteoarthritis and musculoskeletal diseases (ESCEO) an. *Osteoporos Int* 2019;30:45–57.
19 Davis JC, Robertson MC, Comans T, *et al*. Guidelines for conducting and reporting economic evaluation of fall prevention strategies. *Osteoporos Int* 2011;22:2449–59.
20 Drummond M, Maetzel A, Gabriel S, *et al*. Towards a reference case for use in future economic evaluations of interventions in osteoarthritis. *J Rheumatol* 2003.
21 Maetzel A, Tugwell P, Boers M, *et al*. Economic evaluation of programs or interventions in the management of rheumatoid arthritis: defining a consensus-based reference case. *J Rheumatol* 2003;30:891–6.
22 Husereau D, Drummond M, Augustovski F, *et al*. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. *Int J Technol Assess Health Care* 2022;38:e13.
23 Brook RH. The RAND/UCLA appropriateness method. In: *Clinical practice guideline development: methodology perspectives*. 1994.
24 Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–15.
25 Liberati A, Altman DG, Tetzlaff J, *et al*. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009;62:e1–34.
26 Page MJ, Moher D. Evaluations of the uptake and impact of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and extensions: A Scoping review. *Syst Rev* 2017;6:263.
27 Diamond IR, Grant RC, Feldman BM, *et al*. Defining consensus: a systematic review recommends Methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;67:401–9.
28 Boulkedid R, Abdoul H, Loustau M, *et al*. Using and reporting the Delphi method for selecting Healthcare quality indicators: A systematic review. *PLoS One* 2011;6:e20476.
29 Ramsey S, Willke R, Briggs A, *et al*. Good research practices for cost-effectiveness analysis alongside clinical trials: the ISPOR RCT-CEA task force report. *Value in Health* 2005;8:521–33.
30 Dalziel K, Li J, Scott A, *et al*. Accuracy of patient recall for self-reported doctor visits: is shorter recall better *Health Econ* 2018;27:1684–98.