Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort

Myles, Paul S; Boney, Oliver; Botti, Mari; Cyna, Allan M; Gan, Tongjoo J; Jensen, Mark P; Kehlet, H; Kurz, A; De Oliveira, G S (Jr); Peyton, P; Sessler, D I; Tramer, Martin; Wu, C L

How to cite

MYLES, Paul S et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. In: British Journal of Anaesthesia, 2018, vol. 120, n° 4, p. 705–711. doi: 10.1016/j.bja.2017.12.037

This publication URL: https://archive-ouverte.unige.ch//unige:128600
Publication DOI: 10.1016/j.bja.2017.12.037

BJA

British Journal of Anaesthesia, 120 (4): 705–711 (2018)

doi: 10.1016/j.bja.2017.12.037
Advance Access Publication Date: 2 February 2018
Review Article

Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort

P. S. Myles[1,*], O. Boney[2], M. Botti[3], A. M. Cyna[4,5], T. J. Gan[6], M. P. Jensen[7], H. Kehlet[8], A. Kurz[9], G. S. De Oliveira Jr. [10], P. Peyton[11], D. I. Sessler[12], M. R. Tramèr[13], C. L. Wu[14] and the StEP–COMPAC Group

[1]Department of Anaesthesia and Perioperative Medicine, Alfred Hospital and Monash University, Melbourne, Australia, [2]Surgical Outcomes Research Centre, University College Hospital, London, UK, [3]School of Nursing and Midwifery, Deakin University, Geelong, Australia, [4]Discipline of Acute Care Medicine, University of Adelaide, Adelaide, Australia, [5]Department of Anaesthesia and Pain Management, Nepean Hospital, University of Sydney, Sydney, Australia, [6]Department of Anesthesiology, Stony Brook University, Stony Brook, NY, USA, [7]Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA, [8]Section for Surgical Pathophysiology, Rigshospitalet, Copenhagen, Denmark, [9]Department of Anesthesiology, Cleveland Clinic, Cleveland, OH, USA, [10]Department of Anesthesiology, Center for Healthcare Studies, Institute for Public Health and Medicine, Northwestern University, Chicago, IL, USA, [11]Anaesthesia, Perioperative and Pain Medicine Unit, Melbourne Medical School, University of Melbourne, Melbourne, Victoria, Australia, [12]Department of Outcomes Research, Anesthesiology Institute, Cleveland Clinic, Cleveland, OH, USA, [13]Department of Anaesthesiology, Geneva University Hospitals, Geneva, Switzerland, [14]Department of Anesthesiology, The Johns Hopkins Hospital, Baltimore, MD, USA

*Corresponding author. E-mail: p.myles@alfred.org.au.

StEP Steering Committee members: Paul Myles (Australia) and Michael Grocott (UK), Co-Chairs; Bruce Biccard (South Africa), Jane Blazeby (UK), Oliver Boney (UK), Matthew Chan (Hong Kong), Elisabeth Diouf (Senegal), Lee Fleisher (USA), Cor Kalkman (The Netherlands), Andrea Kurz (USA), Ramani Moonesinghe (UK), and Duminda Wijeysundera (Canada).

Abstract

Background: Maximising patient comfort during and after surgery is a primary concern of anaesthetists and other perioperative clinicians, but objective measures of what constitutes patient comfort in the perioperative period remain poorly defined. The Standardised Endpoints in Perioperative Medicine initiative was established to derive a set of standardised endpoints for use in perioperative clinical trials.

Methods: We undertook a systematic review to identify measures of patient comfort used in the anaesthetic, surgical, and other perioperative literature. A multi-round Delphi consensus process that included up to 89 clinician researchers was then used to refine a recommended list of outcome measures.

Editorial decision: December 29, 2017; **Accepted:** December 29, 2017

Results: We identified 122 studies in a literature search, which were the basis for a preliminary list of 24 outcome measures and their definitions. The response rates for Delphi Rounds 1, 2, and 3 were 100% (n=22), 90% (n=79), and 100% (n=13), respectively. A final list of six defined endpoints was identified: pain intensity (at rest and during movement) at 24 h postoperatively, nausea and vomiting (0–6 h, 6–24 h, and overall), one of two quality-of-recovery (QoR) scales (QoR score or QoR-15), time to gastrointestinal recovery, time to mobilisation, and sleep quality.
Conclusions: As standardised outcomes will support benchmarking and pooling (meta-analysis) of trials, one or more of these recommended endpoints should be considered for inclusion in clinical trials assessing patient comfort and pain after surgery.

Keywords: anaesthesia; clinical trials; patient-reported outcomes; surgery

Editor's key points

- Objective measures of patient comfort in the perioperative period remain poorly defined.
- A systematic literature review and Delphi consensus process were used to identify and define recommended outcome measures.
- A final list of six defined endpoints was identified, including pain, postoperative nausea and vomiting, quality-of-recovery scales, gastrointestinal recovery, mobilisation, and sleep quality.
- These recommended endpoints will facilitate future clinical trials assessing patient comfort and pain after surgery.

Maximising patient comfort during and after surgery is a primary concern of anaesthetists and other perioperative clinicians. However, the objective measures of patient comfort in the perioperative period remain poorly defined. There is a pressing need to standardise endpoints in perioperative clinical trials so that results can be compared across studies.[1]

Recent studies have suggested various measures,[2–4] but these recommendations were not developed in accord with current standards for the development of guidelines or consensus statements.[5–7] Consensus and consistency in the use of appropriate perioperative outcome measures, and their timing of assessments, would enhance the interpretation and translation of patient-centred outcomes research and improve the validity of data synthesis in meta-analysis. The use of standardised endpoints will presumably improve the ability of clinicians to compare results across trials, and for the investigators to pool results from diverse studies.

The overall aim of the Standardised Endpoints in Perioperative Medicine (StEP) initiative is to derive a set of endpoints for use in perioperative-medicine trials, based on current evidence, expert guidance, and international consensus.[1] Here, we describe the results of a systematic review and Delphi process to identify important measures of patient comfort in the perioperative setting. Patient satisfaction, quality of life, and other health-status scales are being evaluated and will be reported separately.[1]

Methods

We first undertook a systematic review to identify measures of patient comfort in the anaesthetic, surgical, and other perioperative literature. A Delphi consensus process was then used to refine a recommended list of outcome measures.

We defined the perioperative period as the pre-, intra-, and post-operative phases of a patient's surgical journey, ranging from preoperative evaluation and planning to expected full recovery. The target population was any patient having any surgical procedure requiring anaesthesia care. Patient comfort measures were defined as any subjective or objective measure of either physical, mental, or emotional comfort and well-being; discomfort or absence of comfort; or symptoms that could be described as undesirable.

Literature search

Systematic searches of MEDLINE, Embase, Web of Knowledge, and the Cochrane database were performed to identify systematic reviews of trials reporting patient-comfort or pain-relief measures, and published guidelines and consensus statements or recommendations regarding the measurement of perioperative patient comfort or bothersome pain. This process, therefore, collated all relevant endpoints that were used in any of the original perioperative trials. Only reviews published from January 1, 2000 onwards were included to reduce the risk of retrieving outdated or obsolete measures. Two authors (O.B. and P.S.M.) independently identified titles and abstracts of potentially eligible studies. Discrepancies were resolved by consultation between the initial two researchers. Where consensus opinion could not be reached, a third researcher was to be consulted, but adjudication was never needed. Reference lists of relevant clinical studies and review articles were explored for additional studies. A detailed description of the MEDLINE search strategy is presented in the Supplementary Table S1.

The intended population included adults (age ≥18 yr) having any type of surgery. Exclusion criteria were any systematic review including patients <18 yr, any review where an operation (using a surgical incision) was not performed (e.g. endoscopy), and any review including data from fewer than 100 patients in the analysis. Additionally, we restricted our search to the four highest impact-factor-ranked specialty journals in anaesthesia (*British Journal of Anaesthesia*, *Anesthesiology*, *Anesthesia & Analgesia*, and *Anaesthesia*), surgery (*Annals of Surgery*, *British Journal of Surgery*, *Journal of the American College of Surgeons*, and *JAMA Surgery*), and general medicine (*New England Journal of Medicine*, *The Lancet*, *The Journal of the American Medical Association*, and the *British Medical Journal*) in May 2016. Guidelines and reviews published in other journals, but identified from hand searching of reference lists of included studies were also included if they met the other study inclusion criteria.

A total of 1044 potential studies were retrieved, of which 102 met the inclusion criteria. Hand searching of reference

lists of included studies retrieved a further 233 potentially relevant reviews and consensus guidelines, of which 20 also met the inclusion criteria. One hundred and twenty-two studies were, therefore, included in this review (Fig. 1). The results of this systematic review informed the following Delphi consensus process.

Delphi process

A Delphi process was used to obtain input and consensus from a group of medical and other health researchers with experience in anaesthesia and perioperative-medicine trials.[8,9] The StEP Working Group consisted of experienced perioperative triallists and other investigators from various countries (see Supplementary Data) and was overseen by a Steering Committee (see Appendix).[1,10]

The filtering and development of recommended perioperative clinical-trial endpoints were undertaken in four stages.[11]

Stage 1: establishing a preliminary list of trial endpoints and their definitions

Publications identified in the literature search were used to create a preliminary list of outcome measures and their definitions (Fig. 1, and Supplementary Tables S2 and S3).

Stage 2: formal rating of the recommendations (Delphi Round 1)

The list of proposed endpoints and their definitions was sent to all members of the StEP theme subgroup (n=13) and the StEP Steering Committee (n=9), using a Delphi questionnaire (Supplementary Table S4). The participants were asked to score each of the items listed using a scale of 1–9, with 1–3 labelled 'not that important or invalid', 4–6 labelled 'important but requires revision', and 7–9 labelled 'critical for inclusion'.[11,12] The participants were given the option to select 'unsure' if they were unable to offer an opinion as to which category to apply to the item. The participants were invited to suggest any other endpoints, or definitions, or modifications to existing definitions that they believed should be added when identifying endpoints in the planning of clinical trials addressing perioperative patient comfort. A reminder e-mail was sent to ensure prompt completion of the survey. The final numbers of respondents and item completions were recorded.

Stage 3: Delphi Round 2

The second Delphi round was coordinated by the StEP theme group Chair (P.M.) and the participants were broadened to include the entire StEP Working Group (n=89). For each item in the first Delphi round, the number of participants who scored

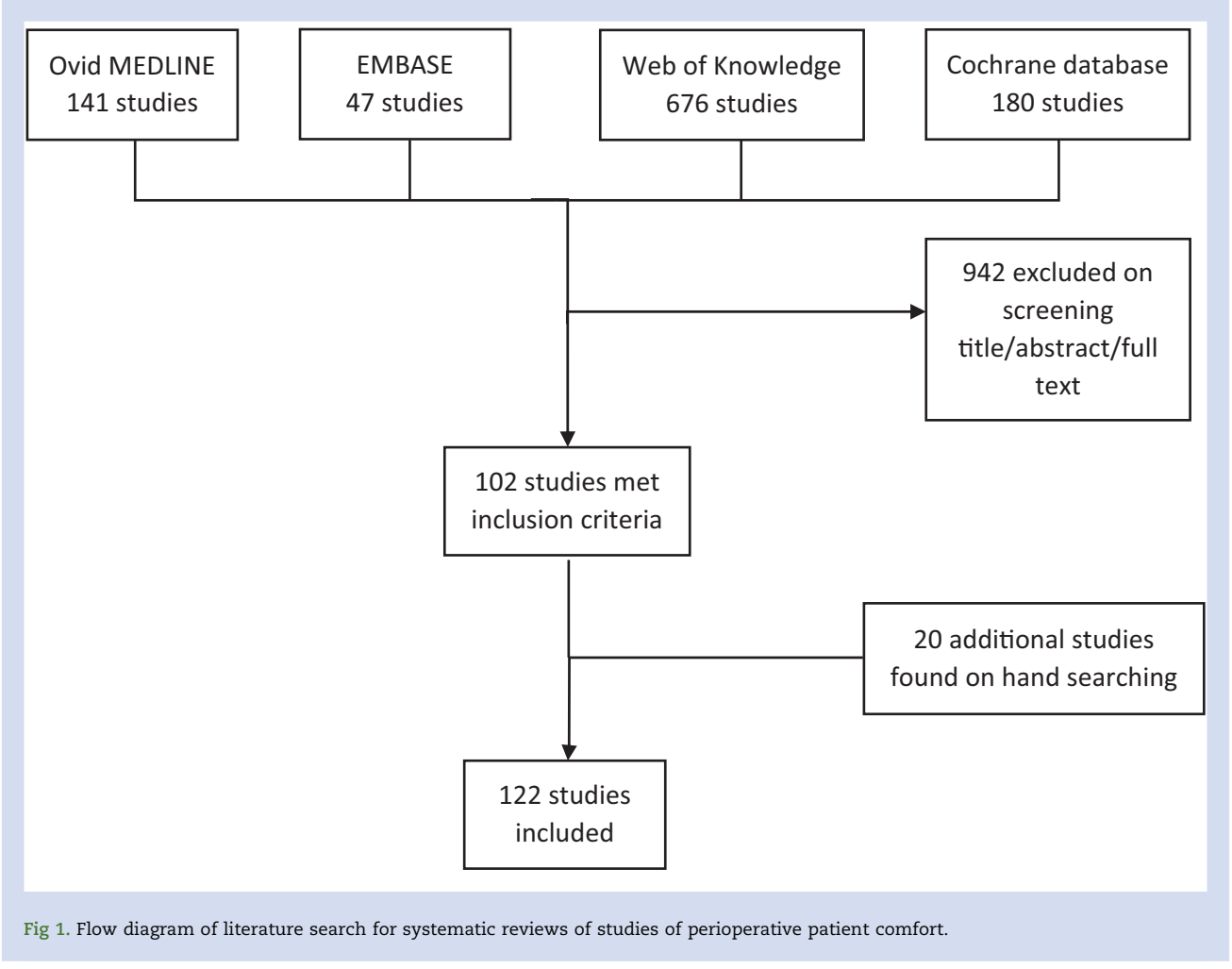

Fig 1. Flow diagram of literature search for systematic reviews of studies of perioperative patient comfort.

the item and the median and inter-quartile range of scores were quantified. Members of the patient-comfort theme group were then invited to discuss the results via e-mail. Any items not rated as critical (i.e. 70th centile score <7), but still with a median score of ≥7 were retained for consideration in the second round.[8] Lower-rated endpoint items identified for removal could be retained if they were considered as critical by any group member for the second round; items with a median score of ≤3 were not retained. Additional endpoints or modified definition items were reviewed by the theme group. Merging or splitting some endpoints was considered at this time.

A reduced set of items was thus carried forward to Delphi Round 2, using the previous proforma. Quality-of-recovery (QoR) scales were then separately evaluated. The participants were provided with the number (%) of respondents scoring ≥7 for each item in Round 1. They were asked to consider the responses from other Delphi participants, and to re-score the item using a second, modified questionnaire (Supplementary Table S5). A consensus was defined by the support from at least 70% of participants scoring 'critical' (score ≥7).[8,11]

Stage 4: developing final recommendations and Delphi Round 3

Delphi Round 3 was coordinated by the StEP theme group Chair (P.M.) and the participants were restricted to the authors, who were all members of the StEP patient-comfort group (n=13). The summary results of the aforementioned process were provided to these participants, inviting further comments. If responses to this final stage suggested a need for further modification to endpoint definitions, then this was resolved by the authors via e-mail discussion. Any endpoint items not rated as critical for inclusion (median score ≤7) were not retained for this round. Each endpoint was additionally rated for validity—specifically, face validity (if this endpoint actually measure the outcome of interest) or content validity (if this endpoint reflect the patient outcome of interest); reliability (if the endpoint is reproducible; feasibility (if the endpoint data can be collected by research staff with some training, without undue effort or risk of missing data; and patient centredness [if the endpoint has a meaningful impact on a patient's recovery (any of discomfort or distress, prolonged hospital stay, need for re-operation, ongoing disability, or increased risk of death)].

Despite not achieving at least 70% support in the second Delphi round, sleep disturbance was identified as a likely important endpoint by some of the authors, and alternative sleep scales were suggested. These included the Pittsburgh Sleep Quality Index,[13] the Patient-Reported Outcomes Measurement Information System (PROMIS) sleep disturbance item bank,[14] and the Richards–Campbell Sleep Questionnaire.[15]

Each Delphi round was coordinated by the Health Services Research Centre of the Royal College of Anaesthetists in the United Kingdom, or the Research Unit of the Department of Anaesthesia and Perioperative Medicine at the Alfred Hospital in Melbourne, Australia. The item scores and number of respondents were recorded for each of the Delphi rounds in an Excel spreadsheet (Microsoft Corporation, Redmond, WA, USA), and then converted to an SPSS (V23.0, SPSS Inc., Chicago, IL, USA) database for calculation of the final median and range scores, and consensus rates.

Results

A long list of potentially relevant patient-comfort outcome measures is presented in Supplementary Table S2. The response rates for Delphi Rounds 1, 2, and 3 were 100% (n=22), 90% (n=79), and 100% (n=13), respectively. The results of the Delphi rounds are presented in Table 1.

Seven of the 22 items nominated in the first Delphi round were removed for the second Delphi round [Item 1 (supplementary analgesic use) was retained following a request from one respondent], and another (Opioid-Related Symptom Distress Scale) was on the suggestion of another respondent. Highly rated patient-comfort outcome measures were carried forward to the second Delphi round (Supplementary Table S5), and the four QoR scales were presented separately. Sixteen of the 23 items nominated in the second Delphi round were removed for the third Delphi round. The supplementary round used to identify the optimal sleep scale resulted in support of the PROMIS sleep disturbance item bank. (Further details are available in Supplementary Table S6.) This resulted in six endpoints, with QoR evaluated with either of two scales (Table 2): (1) postoperative pain intensity at rest and on movement at 24 h, using a numerical rating scale [0–10, with 0='no pain' and 10='insert maximum endpoint descriptor here (e.g. worst pain possible)'], and reported for pain at 24 h and (ideally) at least one other time point; (2) incidence of postoperative nausea, vomiting/retching, and nausea and vomiting [early (defined as 0–6 h), late (defined as 6–24 h), and overall (use of rescue antiemetic), each reported as number (%)][16]; (3) postoperative QoR using either the nine-item QoR score[17] or the 15-item QoR-15 scale[18,19]; (4) time to gastrointestinal recovery, defined by time to tolerate oral diet (e.g. soft food or light meal); and (5) time to mobilisation; and sixth, sleep disturbance, using the PROMIS scale.[14]

Discussion

We undertook a systematic review and used a Delphi process to achieve consensus from a broad range of experts involved in perioperative clinical studies. The group identified six standardised endpoints that represent key aspects of patient comfort in the perioperative setting. We recommend that one or more of these endpoint domains be considered when assessing patient comfort in perioperative trials.

There have been previous attempts at identifying outcome measures after surgery, anaesthesia, and critical care.[2–4,20] However, we included a broader representation of medical and nursing disciplines, from several countries with differing health systems, and used a Delphi methodology to achieve consensus.[5] We also considered the validity, reliability, feasibility, and patient centredness of these endpoints.[6,7] This resulted in a smaller set of endpoints, each with a clear definition and, where available, reference to supporting validation studies. The timing of measurement of any aspect of patient comfort will influence the result, but we have chosen to not stipulate timing for most endpoints because these should be guided by the expected effects of a study intervention.

Clinical researchers do not always measure what is important to patients.[21–23] Patients value access to clear information; a sense of comfort before, during, and after surgery; and a speedy recovery and early return to their own home.[17,23,24] Several aspects of postoperative pain and comfort that some readers might consider important were considered, but rejected after the Delphi process, including

Table 1 Results of the Delphi rounds. ncf, not carried forward; nd, not done; PONV, postoperative nausea and vomiting; PQRS, postoperative quality recovery scale; QoR, quality of recovery

Summary of item*	Delphi Round 1 (*n*=22)			Delphi Round 2 (*n*=79)			Delphi Round 3 (*n*=12)		
	Unsure (*n*)	Median score	Scores ≥7 (%)	Unsure (*n*)	Median score	Scores ≥7 (%)	Unsure (*n*)	Median score	Scores ≥7 (%)
1. Supplementary analgesic use	0	6	45	3	5	27	—	ncf	—
2a. Preoperative anxiety	1	5	38	—	ncf	—	—	ncf	—
2b. Perioperative anxiety	3	5	37	—	ncf	—	—	ncf	—
3a. Morphine consumption at 6 h	1	5	33	—	ncf	—	—	ncf	—
3b. Morphine consumption at 24 h	0	7	68	1	7	54	—	ncf	—
3c. Morphine consumption at 48 h	0	7	54	2	6	40	—	ncf	—
4. Subjective analgesic effectiveness	0	7	54	2	6	44	—	ncf	—
5a. Pain intensity (rest, movement) at 6 h	1	7	81	4	8	78	—	ncf	—
5b. Pain intensity (rest, movement) at 12 h	0	7	59	1	7	61	—	ncf	—
5c. Pain intensity (rest, movement) at 24 h	0	8	86	3	8	92	0	9	92
5d. Pain intensity (rest, movement) at 72 h	1	8	76	2	7	68	—	ncf	—
6. Time to first analgesic request	0	5	41	—	ncf	—	—	ncf	—
7. Opioid-related side-effects	0	7	68	2	7	68	—	ncf	—
7a. Opioid-Related Symptom Distress Scale	—	nd	—	4	7	51	—	ncf	—
8. Incidence of PONV (0–6 h, 6–24 h, and overall)	0	7	64	1	7	87	0	8	92
9. Severity of PONV	0	7	59	4	7	53	—	ncf	—
10. Incidence of post-discharge nausea, vomiting	0	6	50	2	7	64	—	ncf	—
11. Severe PONV at 24 h	0	7	59	1	7	58	—	ncf	—
12. Antiemetic complete response	1	7	52	1	6	29	—	ncf	—
12. Need for rescue antiemetics	1	6	42	—	ncf	—	—	ncf	—
13. QoR scales	2	9	90	—	—	—	—	—	—
QoR score		(any of)	(any of)	9	8	86	1	8	100
QoR-40				8	7	51	—	ncf	—
QoR-15				9	8	85	1	8	100
14. QoR (PQRS scale)	6	7	62	11	7	55	—	ncf	—
15. Time to gastrointestinal recovery	0	7	82	1	7	77	1	8	92
16. Time to mobilisation	0	8	77	1	8	87	0	8	85
17. Shivering	0	4	14	—	ncf	—	—	ncf	—
18. Perioperative thermal comfort	3	5	21	—	ncf	—	—	ncf	—
19. Joint arthroplasty range of motion at 24 h	2	6	35	—	ncf	—	—	ncf	—
20. Fatigue	2	6	40	—	ncf	—	—	ncf	—
21. Sleep disturbance	2	7	70	4	7	57	1	8	83
22. Adverse events	0	7	54	0	6	39	—	ncf	—

* See Methods section and the Supplementary Tables for complete definitions.

opioid consumption and opioid-related side-effects, additional pain intensity scoring, and patient anxiety and well-being. Some components of these are included in the QoR scales, or will manifest as delayed recovery or sleep disturbance, and so their impact can be evaluated using these tools. Although this study focused on patient comfort, there are other related aspects of recovery, including well-being, functional status, and quality of life. Loss of the ability to live independently is a particular concern for the elderly.[25,26] These are being addressed by other StEP groups (see Supplementary Data).[1]

Patient comfort can be influenced by physical symptoms and distress, or psychological or spiritual concerns. External factors may also play a role, including patients' physical and social environment, access to care, and healthcare costs. Patient comfort will be affected by anxiety and other emotional factors, and physical discomforts, such as wound pain, thirst, sore throat, dizziness, headache, constipation, urinary retention, and other symptoms commonly experienced after surgery and anaesthesia. Several QoR scales have been developed to encapsulate these outcomes, but few have been extensively validated in a broad range of surgical settings. The QoR-40 has been extensively validated and previously recommended,[27–31] but the QoR score (nine items) and QoR-15 scales, which have also been externally validated[32,33] and are shorter, were recommended by our Delphi process. Both the QoR score and QoR-15 scales provide patient-centred global assessments of patients' postoperative recovery.

Our process has some important limitations. For example, we acknowledge that the patients' experience of comfort should be evaluated from their perspective, and so our decision to not include a patient representative in this project deserves scrutiny. However, our purpose was to identify valid and robust endpoints currently used in clinical trials. We, therefore, restricted our analysis to information derived from systematic reviews and meta-analyses in relevant leading journals and to researchers representing anaesthesia, critical care, surgery, and nursing. We fully accept that there are other perspectives that might be of value, and we plan to incorporate patient representatives in future work focused on identifying a core outcome set for perioperative medicine and surgery.[1] In addition, many of the endpoints that were identified for evaluation in this project had been developed with direct patient involvement.[14,17] Interestingly, few authors have explicitly reported patient comfort or bothersome pain.[34,35] These and other endpoints that were either not considered or discarded during the Delphi process may

Table 2 Results of the final Delphi round: psychometric and clinical assessments (n=13). PONV, postoperative nausea and vomiting; QoR, quality of recovery. *Validity = Does the endpoint and its definition have face validity (in your opinion, this endpoint actually measures the outcome of interest) or content validity (this endpoint reflects the patient outcome of interest)? †Reliability = Is the endpoint reproducible (if the endpoint was collected by others in similar settings)? ‡Feasibility = Can the endpoint data be collected by research staff with some training, without undue effort or risk of missing data? ¶Patient centredness = Does the endpoint have a meaningful impact on a patient's recovery (any of the following: discomfort or distress, prolonged hospital stay, ongoing disability, or increased risk of death)?

Item*	Validity*			Reliability†			Feasibility‡			Patient centredness¶		
	Unsure (n)	Median score	Scores ≥7 (%)	Unsure (n)	Median score	Scores ≥7 (%)	Unsure (n)	Median score	Scores ≥7 (%)	Unsure (n)	Median score	Scores ≥7 (%)
Pain intensity (rest and movement) at 24 h	0	9	92	0	7	92	0	9	100	0	9	92
Incidence of PONV (0–6 h, 6–24 h, and overall)	2	8	91	1	8	92	0	9	92	0	9	85
QoR scales												
QoR score	1	8	100	1	8	100	1	8	100	1	8	100
QoR-15	1	8	100	1	8	100	1	8	100	1	8	100
Time to gastrointestinal recovery	1	8	92	1	7	83	1	8	100	1	8	92
Time to mobilisation	0	8	85	0	8	85	0	8	100	0	8	92
Sleep disturbance	1	8	75	1	8	69	1	8	75	1	8	75

eventually be found to be valid and patient centred. We encourage such studies and expect to update these StEP recommendations in the future.

In conclusion, we identified six key patient-comfort outcome measures that should be considered by those designing perioperative clinical studies. Standardised endpoints will facilitate uniform data reporting, and will support improved benchmarking and meta-analyses of future trials.

Authors' contributions

Study concept: P.S.M.
Protocol development: P.S.M.
Systematic review and assistance with protocol development: O.B.
Participation in the Delphi survey: all authors.
Analysis of the responses: all authors.
First draft and revision of manuscript: P.S.M.
Critical review and revisions of the manuscript: all authors.

Acknowledgements

The authors thank Ms Laura Farmer and James Goodwin of the National Institute of Academic Anaesthesia, Health Services Research Centre (Royal College of Anaesthetists, London, UK), and Ms Angela Lowe (Alfred Hospital, Melbourne, Australia) for their assistance with the Delphi surveys, and all members of the StEP–COMPAC Group (listed in the Supplementary Data) who responded in a timely fashion. The authors thank the *British Journal of Anaesthesia* for sponsoring a workshop at the Monash University Prato Centre, Italy, in June 2015.

Declaration of interest

P.S.M. is an editor of the *British Journal of Anaesthesia*, and the developer of the quality-of-recovery scales. None of the other authors report any competing interests relating to the topic of this paper.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.bja.2017.12.037.

References

1. Myles PS, Grocott MP, Boney O, Moonesinghe SR. Standardizing end points in perioperative trials: towards a core and extended outcome set. *Br J Anaesth* 2016; **116**: 586–9
2. Jammer I, Wickboldt N, Sander M, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: european Perioperative Clinical Outcome (EPCO) definitions: a statement from the ESA–ESICM joint taskforce on perioperative outcome measures. *Eur J Anaesthesiol* 2015; **32**: 88–105
3. Stundner O, Ortmaier R, Memtsoudis SG. Which outcomes related to regional anesthesia are most important for orthopedic surgery patients? *Anesthesiol Clin* 2014; **32**: 809–21
4. Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess* 2001; **5**: 1–194

5. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med* 2010; **7**, e1000217
6. Revicki DA, Gnanasakthy A, Weinfurt K. Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the PRO Evidence Dossier. *Qual Life Res* 2007; **16**: 717–23
7. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil* 2000; **81**: S21–9
8. 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
9. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS Med* 2011; **8**, e1000393
10. Boney O, Moonesinghe SR, Myles PS, Grocott MP. Standardizing endpoints in perioperative research. *Can J Anaesth* 2016; **63**: 159–68
11. Kirkham JJ, Gorst S, Altman DG, et al. COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). *Trials* 2015; **16**: 373
12. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 2. framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011; **64**: 395–400
13. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The Pittsburgh Sleep Quality Index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. *Sleep Med Rev* 2016; **25**: 52–73
14. Pilkonis PA, Yu L, Colditz J, et al. Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS): use, consequences, and expectancies. *Drug Alcohol Depend* 2013; **130**: 167–77
15. Richards KC, O'Sullivan PS, Phillips RL. Measurement of sleep in critically ill patients. *J Nurs Meas* 2000; **8**: 131–44
16. Gan TJ, Diemunsch P, Habib AS, et al. Consensus guidelines for the management of postoperative nausea and vomiting. *Anesth Analg* 2014; **118**: 85–113
17. Myles PS, Hunt JO, Nightingale CE, et al. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesth Analg* 1999; **88**: 83–90
18. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. *Anesthesiology* 2013; **118**: 1332–40
19. Myles PS, Myles DB, Galagher W, Chew C, MacDonald N, Dennis A. Minimal clinically important difference for three quality of recovery scales. *Anesthesiology* 2016; **125**: 39–45
20. Macefield RC, Boulind CE, Blazeby JM. Selecting and measuring optimal outcomes for randomised controlled trials in surgery. *Langenbecks Arch Surg* 2014; **399**: 263–72
21. Pluut B. Differences that matter: developing critical insights into discourses of patient-centeredness. *Med Health Care Philos* 2016; **19**: 501–15
22. Griffiths F, Mason V, Boardman F, et al. Evaluating recovery following hip fracture: a qualitative interview study of what is important to patients. *Br Med J Open* 2015; **5**, e005406
23. Lavallee DC, Chenok KE, Love RM, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff (Millwood)* 2016; **35**: 575–82
24. Beaudreau SA, Spira AP, Stewart A, et al. Validation of the Pittsburgh sleep quality Index and the epworth sleepiness scale in older black and white women. *Sleep Med* 2012; **13**: 36–42
25. Berian JR, Mohanty S, Ko CY, Rosenthal RA, Robinson TN. Association of loss of independence with readmission and death after discharge in older patients after surgical procedures. *JAMA Surg* 2016; **151**, e161689
26. Fried TR, Bradley EH, Towle VR, Allore H. Understanding the treatment preferences of seriously ill patients. *N Engl J Med* 2002; **346**: 1061–6
27. Kluivers KB, Riphagen I, Vierhout ME, Brolmann HA, de Vet HC. Systematic review on recovery specific quality-of-life instruments. *Surgery* 2008; **143**: 206–15
28. Herrera FJ, Wong J, Chung F. A systematic review of postoperative recovery outcomes measurements after ambulatory surgery. *Anesth Analg* 2007; **105**: 63–9
29. Gornall BF, Myles PS, Smith CL, et al. Measurement of quality of recovery using the QoR-40: a quantitative systematic review. *Br J Anaesth* 2013; **111**: 161–9
30. Yaghoobi S, Hamidfar M, Lawson DM, Fridlund B, Myles PS, Pakpour AH. Validity and reliability of the Iranian version of the quality of recovery-40 questionnaire. *Anesth Pain Med* 2015; **5**, e20350
31. Tanaka Y, Wakita T, Fukuhara S, et al. Validation of the Japanese version of the quality of recovery score QoR-40. *J Anesth* 2011; **25**: 509–15
32. Chazapis M, Walker EM, Rooms MA, Kamming D, Moonesinghe SR. Measuring quality of recovery-15 after day case surgery. *Br J Anaesth* 2016; **116**: 241–8
33. Chan MT, Lo CC, Lok CK, Chan TW, Choi KC, Gin T. Psychometric testing of the Chinese quality of recovery score. *Anesth Analg* 2008; **107**: 1189–95
34. Chooi CS, White AM, Tan SG, Dowling K, Cyna AM. Pain *vs* comfort scores after Caesarean section: a randomized trial. *Br J Anaesth* 2013; **110**: 780–7
35. Chooi CS, Nerlekar R, Raju A, Cyna AM. The effects of positive or negative words when assessing postoperative pain. *Anaesth Intensive Care* 2011; **39**: 101–6

Handling editor: H.C. Hemmings Jr