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

S. Ramani Moonesinghe,*[1 2 3] Alexander I.R. Jackson,[4 5] Oliver Boney,[1 2] Nathalie Stevenson [6] Matthew T.V. Chan, [7] Tim M. Cook[8 9], Meghan Lane-Fall,[10 11] Cor Kalkman, [12] Mark D. Neuman,[13] Ulrica Nilsson [14 15], Mark Shulman, [16 17] Paul S Myles [16 17] and the StEP-COMPAC group.

* Corresponding author: S. Ramani Moonesinghe. @rmoonesinghe

[1] UCL/UCLH Surgical Outcomes Research Centre, Centre for Perioperative Medicine, Department for Targeted Intervention, University College London, London, UK
[2] Health Services Research Centre, National Institute for Academic Anaesthesia, Royal College of Anaesthetists, London, UK.
[3] National Institute for Health Research University College London Hospitals Biomedical Research Centre, London, UK.
[4] Anaesthesia and Critical Care Research Area, NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK
[5] Integrative Physiology and Critical Illness Group, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK
[6] Department of Anaesthesia and Intensive Care Medicine, Barnet General Hospital, Royal Free London NHS Foundation Trust, London, UK
[7] Department of Anaesthesia and Intensive Care, The Chinese University of Hong Kong, Hong Kong, Special Administrative Region, China
[8] Department of Anaesthesia and Intensive Care Medicine, Royal United Hospital Bath NHS Foundation Trust, Bath, UK
[9] School of Medicine, University of Bristol, Bristol UK
[10] Department of Anesthesiology and Critical Care, University of Pennsylvania, Philadephia, Pennsylvania, USA
[11] Penn Center for Perioperative Outcomes Research and Transformation, University of Pennsylvania, Perelman School of Medicine
[12] Division of Anesthesiology, Intensive Care and Emergency Medicine, University Medical Centre, Utrecht, The Netherlands.
[13] Leonard David Institute of Health Economics, University of Pennsylvania
[14] Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden
[15] Perioperative Medicine and Intensive Care, Karolinska University Hospital, Stockholm, Sweden
[16] Department of Anaesthesiology and Perioperative Medicine, The Alfred Hospital, Melbourne, Victoria, Australia
[17] Department of Anaesthesia and Perioperative Medicine, Central Clinical School, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia

Abstract

Background

Patient-centred outcomes are increasingly used in perioperative clinical trials. The Standardised Endpoints initiative aims to define which measures should be used in future research to facilitate comparison between studies, and to enable robust evidence synthesis.

Methods

A systematic review was conducted to create a long-list of patient satisfaction, health-related quality of life, functional status, patient wellbeing and 'life impact' measures for consideration. A 3-stage Delphi consensus process involving 89 international experts was then conducted in order to refine this list into a set of recommendations.

Results

The literature review yielded six patient satisfaction measures, seven generic health-related quality of life measures, eight patient wellbeing measures, five functional status measures, and five 'life impact' measures for consideration. The Delphi response rates were 92%, 87% and 100% for rounds 1, 2, and 3 respectively. Three additional measures were added during the Delphi process as a result of contributions from StEP group members. Firm recommendations have been made about one health-related quality of life measure (Euro-QOL 5-Dimension, 5-Level with Visual Analogue Scale), one functional status measure (World Health Organisation-Disability Assessment Schedule v2.0 12-question version), and one life impact measure (days alive and out of hospital at 30 days after surgery). Recommendations with caveats have been made about the Bauer Patient satisfaction measure, and two life impact measures (days alive and out of hospital at 1 year after surgery, and discharge destination).

Conclusion

Several patient-centred outcome measures have been recommended for use in future perioperative studies. We suggest that every clinical study should consider using at least one patient-centred outcome within a suite of endpoints.

Introduction

When evaluating novel interventions in research trials, understanding their impact on outcome from a patient perspective is both important, and now increasingly a requirement of major funders and journals. (1) Initiatives such as the Core Outcome Measures for Effectiveness Trials (COMET) programmes use co-production between clinicians and patients to determine the 'vital few' measures which should be used in all clinical trials, whether as primary or secondary endpoints, and which are considered to be important to both of these important stakeholder groups. In parallel with the Core Outcome Measures in Perioperative and Anaesthesia Care (COMPAC) branch of COMET, the Standardised Endpoints in Perioperative Medicine initiative seeks to provide clear definitions for the outcomes which should be used in clinical trials, so as to synchronise reporting, facilitate comparison and enable evidence synthesis. (2) This manuscript describes the work of the StEP-COMPAC theme group for patient-centred outcomes.

Methods

We conducted a series of systematic reviews in order to draw up a long-list of patient-centred outcome measures which have been used in previously published studies in perioperative medicine. We then undertook a 3-stage Delphi consensus process in order to refine this long-list and produce a short-list of recommended measures.

We used the following definitions

Perioperative: the general definition of perioperative care relates to the entire surgical pathway, from the time of contemplation of surgery, through to long-term follow-up. (3) For the purpose of this consensus process, we included only those measures which included an evaluation of outcomes related to surgery and anaesthesia (therefore measures which solely evaluated aspects of early decision-making or preoperative assessment were excluded).

Patient satisfaction measure: we applied the same criteria as a previous comprehensive systematic review of the literature: (4) an instrument that was developed using psychometric techniques, and that consisted of at least two distinct dimensions

Health-related quality of life measure (HRQOL): A measure which is taken before and after an intervention to evaluate the impact of that intervention on health-related QOL.

Measures of patient well-being: Measures reflecting psychosocial health, including depression and anxiety

Measures of functional status: Any endpoint which measures the American Thoracic Society (ATS) definition of functional status: that is, "*the ability to perform normal daily activities required to meet basic needs, fulfil usual roles, and maintain health and well-being. Functional status subsumes related concepts of interest: functional capacity and functional performance*".

Life impact measures: endpoints which reflect impact on specific activities such as work, physical status and domicile.

Literature reviews

Patient satisfaction

The group lead (SRM) had previously published a comprehensive systematic review of patient satisfaction measures used in anaesthesia and this was used as the source for the long-list for this domain. (4) Only measures which reported patient satisfaction in our definition of

perioperative care, and which were relevant to our inclusion criteria, were included – therefore measures relating solely to obstetric, paediatric anaesthesia or preoperative assessment were excluded. Further, where questionnaires had been refined or improved, we only included the most recent version; and finally, we only included questionnaires which included questions about patient outcome (e.g. symptoms) as opposed to only questions about patient experience (e.g. quality of information or facilities).

Health-related quality of life

The databases MEDLINE® and Embase® were searched by one author (N.S.) to identify HRQOL measures which had been used in the perioperative literature. The full search strategy (appendix one) combined variations of the term 'perioperative' with variations of the term 'patient reported outcome' (table 1). The search strategy was developed with the assistance of a librarian and with reference to previously published systematic reviews of outcome measures for perioperative care and anaesthesia. (4) (5) We only considered randomised controlled trials which recruited at least 100 adults (over age 18 years) during the years 2000-2016; no language restrictions were placed.

Focused review of high impact journals

In addition to the broad searches above, one author (O.B.) reviewed all randomised controlled trials, published in 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 British Medical Journal and the Journal of the American Medical Association) between January 2005 and May 2016. The purpose was two-fold: to complement the specific searches detailed above; and to identify relevant endpoints which were missed when using standard search terms. These included endpoints which measured the domains of functional status, patient well-being and patient life impact, as defined above.

Following these initial searches, a long-list of trial endpoints and their definitions was created by combining the results of each systematic review, and grouping them into themed domains. We then undertook a 3-stage Delphi consensus process.

Delphi round 1: The long-list was sent to all members of the StEP Patient-Centred Outcomes sub-group (n=12). These individuals were asked to rate each measure using an ordinal scale (1 to 9) with scores 1 to 3 labelled as 'not that important or valid', 4 to 6 labelled 'important but requires revision', and 7 to 9 labelled 'critical for inclusion'. An additional 'unsure' option was also available. In this round, participants were also able to suggest additional endpoints for inclusion, or modifications to definitions. Participants were also encouraged to record free-text comments.

Measures which reached either a median score of ≥7 or which were graded as ≥7 by at least 70% of participants, were forwarded to Delphi round 2. In addition, any measure not meeting these criteria, but considered critically important (i.e. scored ≥7) by any individual group member was retained for Delphi round 2.

Delphi round 2: All participants from the StEP working group (n=89) were invited to take part. This process was coordinated by the NIAA Health Services Research Centre and group member A.I.R.J. A scoring sheet was circulated electronically which included all measures which met the criteria detailed above. The scoring sheet included key information including the reason for inclusion, and the median score, percentage of respondents who graded each measure ≥7, and any free-text comments from Delphi round 1. Measures which reached either a median score of ≥7, or which had agreement from >70% of participants in Delphi round 2, were forwarded to round 3.

Delphi round 3: Participants in Delphi round 3, were restricted to the authors (the StEP patient centred outcomes theme subgroup, n=12). These participants were first provided with the list of measures due to be considered as a result of round 2. They were then invited to suggest additional endpoints for inclusion. A final shortlist was then compiled which included key information from the previous round, including the median score, percentage of respondents who graded each measure ≥7 and any free-text comments. Participants were asked to again grade each measure using the same ordinal scale, and in addition to rate (high, medium, low or unsure) the validity, reliability, feasibility and patient-centredness of each measure as previously described. (6)

1. Validity – Does the endpoint and its definition have face validity (in your opinion this endpoint actually measures the outcome of interest) and/or content validity (this endpoint reflects the patient outcome of interest)?
2. Reliability - Is the endpoint reproducible (if the endpoint was collected by others in similar settings)?
3. Feasibility - Can the endpoint data be collected by research staff with some training, without undue effort or risk of missing data?
4. Patient-centeredness - Does the endpoint reflect 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)?

Results

Literature searches

Nineteen questionnaires which evaluated patient satisfaction with perioperative care were identified in the previously published systematic review. **(4)** Of these, six met our inclusion criteria and were included in the first Delphi consensus round.

The new MEDLINE® and Embase® search of HRQOL measures yielded 7028 manuscripts of which 2080 were duplicates (see supplementary material for flowchart). Following abstract and then full-text review, 116 randomised trials using HRQOL measures were included; these described seven generic HRQOL measures used before and after surgery, which were then included in the first Delphi Round.

The focused review of high impact journals identified all seven of the generic HRQOL measures, and an additional eight patient wellbeing measures, five functional status measures and five 'life impact' measures.

Delphi consensus process

The Delphi process reduced the number of measures from 33 down to nine over three rounds (table 1). The response rates for the three Delphi rounds were 92%, 87% and 100%

respectively. The final Delphi round, including the validity assessment (tables 2 to 6) enabled us to make firm recommendations of one HRQOL measure and one functional outcome measure, and recommendations with caveats on two 'life impact' measures and one patient satisfaction measure.

Patient satisfaction

One patient satisfaction measure (the Bauer Patient Satisfaction questionnaire) (7) was carried through to Delphi round 2 as a result of high median score; another (Caljouw's questionnaire (8)) was carried through to Delphi round 2 on the basis of a high score from one participant. Only the Bauer questionnaire reached the threshold to be carried through to the final, 3rd, Delphi round. Most Delphi round 3 participants rated the Bauer questionnaire as having high or medium validity in all domains; the greatest concern was expressed about its reliability (3 participants (25%) rated this low, one participant was unsure). In addition, concerns were expressed about the risks of self-administration as opposed to structured interviews to elicit responses – patients were more 'critical' in their description of their outcomes and satisfaction when responding in interviews in the original development and validation paper. (7)

Participants in Delphi round 3 also expressed the need for modification of this measure in free-text comments including:

- the need to improve the Likert scale used to describe postoperative discomfort (none, moderate, severe – it was suggested that 'mild' should be added);
- the need for guidance on the timing of administration – it was suggested that it should be measured within 24h of end of surgery, and maximum 48h in order to avoid recall bias;

Patient well-being

Only one of the eight endpoints which were long-listed after the systematic review stage reached the third Delphi consensus round: "complete recovery (self-assessed) at 30 days". However, in Delphi round 3, participants did not rate it highly, (median score 6); nor did it reach the required level of consensus for us to be able to recommend it for use as a trial endpoint. Free-text comments included that the Quality of Life-15 measure, (9) which has

been recommended by the Patient Comfort Endpoints StEP group, (6) would be an appropriate alternative measure of patient well-being. Other reservations included the lack of a baseline evaluation (*"If the patient was completely fit pre-op, the outcome is likely to be different to someone who had a degree of disability already"),* the likely impact of specific context (e.g. the type of surgery), and the lack of formal validity testing in previous research.

Health-related Quality of Life and Functional Outcome

The EQ5D (5-level version with visual analogue scale) (10) was highly rated throughout the Delphi process and in the validity assessment in the final stage. The Short-Form 12 (SF-12) (11) reached Delphi Round 3, but with lower scores, lower consensus and fewer participants rating its validity as high across the five domains of assessment.

The modified Rankin score (12) (within the functional outcome domain) reached the final Delphi round but in this final round scored poorly in all assessments.

Two measures which had not been identified in the literature reviews (the World Health Organisation Disability Assessment Schedules, 12-item and 36-item versions) were added after the first Delphi Round at the request of two group members. The WHODAS 2.0 12-question measure (13) reached the final stage, was scored highly by participants for importance, and also for validity. This was classified under the HRQOL domain for the entire Delphi consensus process but in the final stage, one participant suggested that was more appropriately categorised as a functional outcome measure. The group agreed with this and therefore, in our final recommendations, we have therefore reassigned the domain for the measure to functional status.

As part of Delphi Round 3, we also discussed the timing of when postoperative measures of HRQOL and functional status should be measured. There was debate between 6 and 12 months, but in the end consensus was reached that 12 months should be the default postoperative time-point, but in some circumstances, 6 months after surgery would be acceptable as an alternate (or additional) time-point.

Life impact

One life impact measure (days alive and out of hospital at day 30 after surgery – DAOH30) was added to Delphi round 3 at the request of a group member, in addition to DAOH at one year (DAOH1Y) which was included from the beginning of the process. Although the median score for both of these measures was above our threshold of 7 in the final Delphi round, neither reached our consensus threshold of 70%. Similarly, discharge destination reached the threshold median score of 7 but without consensus between participants. Group members also expressed uncertainty about the validity of all three of these measures, in particular, reliability and feasibility. Free-text comments indicated that these concerns related predominantly to the potential for uncontrolled variation in factors which might affect whether or not a patient is admitted to hospital or discharged to a care facility, both with respect to individual patient characteristics (e.g. social deprivation, stoicism, level of family support, requirement for short-term rehabilitation with an expectation of later discharge home) and organisational characteristics (availability and structure of social care and community based medical or nursing care, and even definitions of 'hospital' or 'care home' which might differ between geographical regions). The lack of consideration of baseline status (i.e. own home or care home) was also cited as a limitation of all these measures. However, after the conclusion of the Delphi process, two further research papers evaluating the validity of the DAOH30 measure were published in the peer-reviewed literature. (14) (15) The manuscripts were based on large datasets from Canada and Europe and specifically evaluated the relationships between DAOH30 (14) (15) and DAOH90 and DAOH180 (14) and a variety of patient characteristics (comorbidity, type of surgery) and short and longer-term clinical outcomes such as morbidity, mortality. We therefore re-appraised both DAOH30 and DAOH1Y and steering group members subsequently considered that DAOH30 had reached the required threshold for inclusion in our final recommendations.

Discussion

We have undertaken a systematic literature review and a Delphi consensus process of international experts in perioperative research, in order to determine the endpoints which should be used to measure patient-centred outcomes in perioperative clinical trials. We are able to make a firm recommendation of the EQ5D (5-level version with visual analogue scale) as a measure of health-related quality of life and the WHO-Disability Assessment Schedule 2.0 12-question version as a measure of functional outcome: both of these endpoints were rated highly by our consensus process, including assessments of their validity, reliability and patient-centredness. The Bauer questionnaire was the highest scoring measure of short-term outcome and satisfaction. It is recommended with the caveat that it should be refined to expand the Likert-scale which is used to evaluate symptoms (to a 4-point scale of none, mild, moderate and severe) and that it should usually be completed within 24h of the end of surgery. We acknowledge that this would preclude its use in patients unable to complete a questionnaire early in their postoperative course (e.g. patients sedated and ventilated after surgery) and therefore if the surgery was of that magnitude or severity, then a later time-point might be appropriate. Recent evaluations in large heterogeneous patient datasets provided the necessary support for a relatively novel life impact measure, DAOH30 to also be recommended for future use. We are unable to recommend a summary measure of patient well-being for use in clinical trials, although the QoR-15 (9) which has been reviewed by another StEP group (6) is likely to be suitable for this purpose.

This endeavour is part of the Standardised End-Points in Perioperative Medicine initiative which seeks to provide guidance for researchers in their selection of outcome measures to be used in clinical effectiveness trials. Our aim is to reduce variation in the measures used, thereby facilitating easier comparison between studies and more reliable synthesis of research evidence. Patient-centred outcomes are likely to be used more and more frequently in clinical trials, given the recognition that traditionally deployed endpoints (such as death or length of hospital stay) incompletely evaluate the impact of clinical interventions on patients. In particular, advances in the quality of anaesthesia, surgical and critical care, mean that short-term mortality is relatively uncommon (recent estimates of patient or 30-day mortality usually sit around 0.5 to 1% after elective major surgery).(16) (17) Therefore, trials will

increasingly be required to use outcomes which are more common, such as short-term complications or longer-term mortality, health-related quality of life or disability, because these are more meaningful to patients and clinicians. A secondary consideration, is that with more common outcomes, smaller sample sizes are required to be able to detect a clinically and statistically meaningful difference between the interventions being evaluated. This combination of improved patient-centeredness and potentially shorter study duration, thereby accelerating the acquisition and dissemination of clinically meaningful trial data, provides a powerful rationale for researchers to use these types of outcome measures in future research.

To that end, while the selection of endpoints used will depend on the aims and objectives of individual studies, we propose that there are few arguments against using a measure of health-related quality of life or functional outcome as one of a suite of primary and secondary outcomes in most clinical effectiveness trials. Some potential disadvantages of longer-term patient-reported measures such as EQ5D or WHODAS, include demand on resources (time and money) to trials - and potentially also to patients. It is true that when considered in the context of most clinical trials, the additional cost incurred to collect patient-reported data is likely to be relatively low; however, smaller and more resource-constrained studies may yet find this a considerable challenge. Longer-term outcomes might be affected by factors unrelated to the surgery itself; however, in the context of a randomised trial, we would expect the risks of such confounding to be balanced out between different groups. There are also potential issues with response rates and attrition bias relating to patient characteristics such as age, gender, socioeconomic group, domiciliary status and ethnicity; (18) however, studies with reasonable resources and clear follow-up protocols, and in which patients have consented to longer-term engagement, have been able to overcome these issues, even across geographical boundaries. (19) Trials which are more resource constrained, particularly in low and middle income settings, may need to consider how to overcome these risks when planning their analyses, particularly the need for *a priori* defined sensitivity analyses to detect differences in responders and non-responders.

In trials of anaesthesia specific interventions, which are aimed at alleviating symptoms and improving patient satisfaction, the Bauer measure is a simple, and feasible endpoint – a UK

study of over 15,000 patients was conducted at very low cost and achieved high response rates from patients. (20) However, it is important to recognise the limitations of this measure, both in terms of its content validity and the Likert scale it uses; we therefore recommend that it is adapted to a 4-point scale, by addition of the 'mild' rating, and that future researchers carefully consider the need for additional measures depending on the interventions being evaluated. In particular, other composite measures of patient comfort such as the QOR-15, which has been recommended by another StEP consensus group,(6) may provide an additional, or more appropriate option.

Our study has three main limitations. First, as the scope of this group's work was very broad, we used a combination of previously published and newly conducted systematic literature reviews to identify potential measures; there was variation in the design of these initial reviews, and in particular, the patient satisfaction review included any study design, whereas the two other reviews included only randomised trials. Second, both of the new literature reviews were conducted by a single researcher, and therefore there is a risk that some measures may have been missed. However, these reviews were simply the starting point of the process, and we are confident that through the engagement of a large international panel, we were able to identify all relevant candidate endpoints. The final limitation is that we included only clinicians in this endeavour and not patients: particularly when determining patient-centred outcomes, the patient perspective could be considered paramount. In mitigation, we know that the measures which we include in our final recommendations have all been developed with patient involvement, and further, the WHODAS, EQ5D and Bauer questionnaires have all been used on thousands of patients internationally. However, we look forward to the reporting of the parallel Core Outcome Measures in Effectiveness Trials programme in Anaesthesia and Perioperative Care which will determine the core outcomes to be used in clinical trials, and in which co-production with patients is a key element. (2) This will complement and build on the work we have completed here, by supporting future researchers to prioritise measures which patients consider to be the most important.

In conclusion, we have made firm recommendations of HRQOL, life impact and functional outcome measures which may immediately be used by researchers designing clinical trials in perioperative medicine. We have also made recommendations with caveats for short-term

patient satisfaction. We commend our findings to researchers, funders and journal editors, in the expectation that this initiative will reduce variation in the choice of outcomes used in future trials, and therefore improve the robust evaluation of clinical interventions for patient benefit.

Author contributions:

Study concept: P.S.M.

Protocol development: P.S.M. and S.R.M.

Systematic reviews: O.B. and N.S.

Management and analysis of literature reviews: S.R.M.

Management of Delphi rounds 1 and 2: A.I.R.J. and S.R.M.

Management and analysis of Delphi round 3: S.R.M. with contributions from all authors.

Participation in the Delphi survey: all authors.

First draft and revision of manuscript: S.R.M.

Critical review and revisions of the manuscript: all authors.

Acknowledgements:

Sincere thanks to James Goodwin and Laura Farmer at the Health Services Research Centre at the Royal College of Anaesthetists for supporting the administrative work associated with this project.

Funding:

This research received no specific funding. SRM is supported by: The Health Foundation (Improvement Science Fellowship); the National Institute for Health Research University College London Hospitals Biomedical Research Centre; Royal College of Anaesthetists (Director, NIAA Health Services Research Centre). AIRJ was supported by the University of Southampton National Institute of Health Research Academic Clinical Fellowship Scheme.

REFERENCES

1. PCORI Engagement Rubric. PCORI (Patient-Centered Outcomes Research Institute) website. 2014
2. 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;116:586-589.
3. Perioperative Medicine: the pathway to better surgical care. Royal College of Anaesthetists. 2013
4. Barnett SF, Alagar RK, Grocott MP, Giannaris S, Dick JR, Moonesinghe SR. Patient-satisfaction measures in anesthesia: qualitative systematic review. Anesthesiology. 2013;119:452-478.
5. Coulman KD, Abdelrahman T, Owen-Smith A, Andrews RC, Welbourn R, Blazeby JM. Patient-reported outcomes in bariatric surgery: a systematic review of standards of reporting. Obes Rev. 2013;14:707-720.
6. Myles PS, Boney O, Botti M et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: patient comfort. Br J Anaesth. 2018;120:705-711.
7. Bauer M, Bohrer H, Aichele G, Bach A, Martin E. Measuring patient satisfaction with anaesthesia: perioperative questionnaire versus standardised face-to-face interview. Acta Anaesthesiologica Scandinavica. 2001;45:65-72.
8. Caljouw MA, van Beuzekom M, Boer F. Patient's satisfaction with perioperative care: development, validation, and application of a questionnaire. Br J Anaesth. 2008;100:637-644.
9. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. Anesthesiology. 2013;118:1332-1340.
10. Herdman M, Gudex C, Lloyd A et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727-1736.
11. Hays RD, Sherbourne CD, Mazel R. User's Manual for the Medical Outcomes Study (MOS) Core Measures of Health-Related Quality of Life. RAND Monograph report. 1995
12. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale:

implications for stroke clinical trials: a literature review and synthesis. Stroke. 2007;38:1091-1096.

13. Shulman MA, Myles PS, Chan MT, McIlroy DR, Wallace S, Ponsford J. Measurement of disability-free survival after surgery. Anesthesiology. 2015;122:524-536.
14. Jerath A, Austin PC, Wijeysundera DN. Days Alive and Out of Hospital: Validation of a Patient-centered Outcome for Perioperative Medicine. Anesthesiology. 2019;131:84-93.
15. Bell M, Eriksson LI, Svensson T et al. Days at home after surgery: an integrated and efficient outcome measure for clinical trials and quality assurance. Lancet E Clinical Medicine. 2019;11:18-26.
16. International Surgical Outcomes Study Group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. Br J Anaesth. 2016;117:601-609.
17. Abbott TEF, Fowler AJ, Dobbs TD, Harrison EM, Gillies MA, Pearse RM. Frequency of surgical treatment and related hospital procedures in the UK: a national ecological study using hospital episode statistics. Br J Anaesth. 2017;119:249-257.
18. Hutchings A, Neuburger J, Grosse Frie K, Black N, van der Meulen J. Factors associated with non-response in routine use of patient reported outcome measures after elective surgery in England. Health Qual Life Outcomes. 2012;10:34.
19. Myles PS, Bellomo R, Corcoran T et al. Restrictive versus Liberal Fluid Therapy for Major Abdominal Surgery. N Engl J Med. 2018;378:2263-2274.
20. Walker EM, Bell M, Cook TM, Grocott MP, Moonesinghe SR, SNAP-1 Investigator Group. Patient reported outcome of adult perioperative anaesthesia in the United Kingdom: a cross-sectional observational study. Br J Anaesth. 2016;117:758-766.

	Delphi Round 1 (n=11)[1]			Delphi round 2 (n=77)[2]			Delphi round 3 (n=12)[3]		
	Unsure	Median score	% of scores ≥7	Unsure	Median score	% of scores ≥7	Unsure	Median score	% of scores ≥7
Patient Satisfaction									
EVAN-G	2	4	18	-	-	-	-	-	-
Bauer Patient Satisfaction measure **(7)**	2	7	55	1	7	61	0	6	33
Calijow's Leiden Patient Perioperative Satisfaction Questionnaire	2	4	9	0	4	5	-	-	-
NRS Questionnaire	2	6	27	-	-	-	-	-	-
Heidegger's 29-item questionnaire	2	4	9	-	-	-	-	-	-
Heidelberg peri-anaesthetic questionnaire	2	4	18	-	-	-	-	-	-
Patient wellbeing									
Beck Depression Inventory	2	5	36	-	-	-	-	-	-

Depression/mood using Zung depression index	4	3	0	-	-	-	-	-	-
State Trait Anxiety Inventory	3	6	36	4	5	17	-	-	-
Beck Anxiety Inventory	2	3	18	2	7	49	-	-	-
Mood/depression using Hospital Anxiety and Depression Scale	3	7	45	-	-	-	-	-	-
Psychological General Well Being Index	3	5	9	-	-	-	-	-	-
'Complete recovery' (self-assessed) within 30 days of procedure	1	6.5	45	3	7	55	0	6	42
Self-rating of anxiety level, depression, relaxation and inner peace (VAS-10 each)	1	4.5	36	2	4	17	-	-	-
Health – related quality of life									
EQ5D (5-level version with VAS)	2	7	64	6	8	83	0	8	83
SF36	2	6	18	-	-	-	-	-	-
SF12	2	7	55	5	7	66	0	6.5	50
WHO Quality of Life (brief)	2	5	36	6	5	16	-	-	-
Health Utilities Index Mark 3	3	4	9	-	-	-	-	-	-
PROMIS-10	2	5	9	-	-	-	-	-	-

VAS (0-100) for general health status	2	6	36	-	-	-	-	-	-
WHO Disability Assessment Schedule 2.0 12-item	-	-	-	4	8	73	0	8	92
WHO Disability Assessment Schedule 2.0 36-item	-	-	-	7	6	32	-	-	-
Functional Status									
Modified Rankin Scale	2	7	55	4	7	51	0	5	25
Timed Up and Go test	2	6	27	-	-	-	-	-	-
6-minute walk test	2	6	36	-	-	-	-	-	-
ECOG performance status	2	6	36	5	5	25	-	-	-
Karnofsky performance index	2	6	27	-	-	-	-	-	-
Life Impact									
Time to return to work	1	8	82	1	7	61	-	-	-
Discharge destination (i.e. own home vs. care home etc.)	1	8	73	4	8	74	0	7	58
Days off work at particular endpoints after surgery	1	7	55	3	6	45	-	-	-
Number of days alive and out of hospital (at one year)	1	8	91	3	8	83	0	8	67

Number of days alive and out of hospital (at 30 days)	-	-	-	-	-	-	0	7	58
Restricted activity days / bed rest in previous 14 days	1	5.5	36	2	5	16	-	-	-
Time to return to usual activities	1	7	45	1	7	65	-	-	-

Table 1: Summary results of Delphi consensus process.

[1] 12 individuals invited; 11 responded (92%)
[2] 89 individuals invited; 77 responded (87%)
[3] 12 individuals invited; 12 responded (100%)