

QUALITY AND PATIENT SAFETY

Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine initiative: clinical indicators

Guy Haller^{1,2,*}, Sohail Bampoe³, Tim Cook⁴, Lee A. Fleisher⁵, Michael P. W. Grocott^{6,7}, Mark Neuman⁵, David Story⁸, Paul S. Myles^{2,9} on behalf of the StEP-COMPAC Group[†]

¹Department of Anaesthesiology, Geneva University Hospitals, Geneva, Switzerland, ²Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia, ³Centre for Perioperative Medicine, University College London, UK, ⁴Department of Anaesthesia and Intensive Care, Royal United Hospitals Bath NHS Foundation Trust, Bath, UK, ⁵Departments of Anesthesiology and Critical Care and Medicine, Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA, ⁶Integrative Physiology and Critical Illness Group, Clinical and Experimental Sciences, Faculty of Medicine, University Hospital Southampton, Southampton, UK, ⁷Anaesthesia, Perioperative and Critical Care Research Group, Southampton NIHR Biomedical Research Centre, University Hospital Southampton, Southampton, UK, ⁸Centre for Integrated Critical Care, University of Melbourne, Melbourne, Australia and ⁹Department of Anaesthesiology and Perioperative Medicine, Alfred Hospital and Monash University, Melbourne, Australia

*Corresponding author. E-mail: Guy.Haller@hcuge.ch

[†]Full list is given in Supplementary material S1 and in the section “StEP Steering Committee members”.

Abstract

Background: Clinical indicators are powerful tools to quantify the safety and quality of patient care. Their validity is often unclear and definitions extremely heterogeneous. As part of the International Standardised Endpoints in Perioperative Medicine (StEP) initiative, this study aimed to derive a set of standardised and valid clinical outcome indicators for use in perioperative clinical trials.

Methods: We identified clinical indicators via a systematic review of the anaesthesia and perioperative medicine literature (PubMed/OVID, EMBASE, and Cochrane Library). We performed a three-stage Delphi consensus-gaining process that involved 54 clinician–researchers worldwide. Indicators were first shortlisted and the most suitable definitions for evaluation of quality and safety interventions determined. Indicators were then assessed for validity, reliability, feasibility, and clarity.

Results: We identified 167 clinical outcome indicators. Participation in the three Delphi rounds was 100% ($n=13$), 68% ($n=54$), and 85% ($n=6$), respectively. A final list of eight outcome indicators was generated: surgical site infection at 30 days, stroke within 30 days of surgery, death within 30 days of coronary artery bypass grafting, death within 30 days of

Editorial decision: 12 April 2019; Accepted: 12 April 2019

© 2019 British Journal of Anaesthesia. Published by Elsevier Ltd. All rights reserved.
For Permissions, please email: permissions@elsevier.com

surgery, admission to the intensive care unit within 14 days of surgery, readmission to hospital within 30 days of surgery, and length of hospital stay (with or without in-hospital mortality). They were rated by the majority of experts as valid, reliable, easy to use, and clearly defined.

Conclusions: These clinical indicators can be confidently used as endpoints in clinical trials measuring quality, safety, and improvement in perioperative care.

Registration: PROSPERO 2016 CRD42016042102 (http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016042102).

Keywords: clinical trials; clinical indicators; quality improvement; outcome measures; patient safety; perioperative medicine; standardised endpoint

Editor's key points

- The Standardised Endpoints in Perioperative Medicine (StEP) initiative was established to derive standardised endpoints for use in perioperative clinical trials.
- After a systematic review and Delphi consensus process, a set of eight outcome indicators was identified that should be considered in designing future perioperative clinical trials.
- Use and reporting of these endpoints will support improved benchmarking and meta-analysis of future perioperative trials involving patient safety, quality and improvement.

Developed initially in the manufacturing industry, indicators are increasingly used in hospitals to monitor organisational performance or patient management.¹ Clinical indicators are specifically designed to measure the quality and safety of patient care.² They are increasingly used in perioperative medicine to drive improvement initiatives or assess the overall quality of care provided.³

Clinical indicators can measure structure, process, or outcome-related aspects of perioperative care.^{4,5} Structure indicators measure organisational composition and resource utilisation.² Some examples include staffing levels, equipment, and access to facilities (e.g. 24 h access to a fully staffed emergency theatre). Process indicators measure the way care is (or should be) delivered (e.g. prophylactic antibiotics administration within 60 min before start of surgery).^{6,7} Outcome indicators measure patient-related results of care. They occupy a prominent position in perioperative care as the nature of clinical practice and its improvement largely relies on measures of outcome.⁸ For example, only after the demonstration that hospital-acquired infections are reduced with preoperative administration of antibiotics within 60 min can this administration become an evidenced-based patient quality improvement recommendation. As such, an increasingly larger number of outcome indicators have been developed to guide quality improvement initiatives in anaesthesia and perioperative medicine⁹ where they are used as direct measures of the quality and safety of care provided. They can also be used in clinical trials as primary or secondary endpoints of interventions to improve patient-related quality and safety of care within the perioperative setting.

Because there is limited academic interest in clinical indicators and their use as outcome measures, there is a significant lack of standardised definitions for this type of endpoints.

For instance, according to the Australian Council on Healthcare Standards (ACHS), an unplanned admission to the ICU, a popular and validated outcome indicator is 'an unplanned admission to the Intensive Care Unit within 24 hours of a procedure with an anaesthetist in attendance', whereas for the Anaesthesia Quality Institute (USA), it is defined as 'an unplanned admission to the intensive care unit within 48 hours of induction'.^{10,11} In addition, only a limited number of indicators have undergone a formal validation process,¹² resulting in the use of poorly defined and validated measures of outcomes.¹³

There is a need to identify and provide clearly defined, reliable, and validated clinical outcome indicators that can be used as endpoints in both perioperative clinical trials and health services research assessing quality and safety improvement initiatives.¹⁴ The Standardised Endpoints in Perioperative Medicine (StEP) initiative is an international collaboration with the aim of identifying a set of endpoints supported by expert guidance and international consensus for use in perioperative medicine trials. The current study describes the results of a systematic literature review and Delphi process to identify important outcome clinical indicators.

Methods

We used the standard method developed by the University of California and the American Institute of research and development.^{15,16} We undertook a systematic review of the literature followed by a Delphi consensus gathering process in order to refine and validate a list of recommended clinical indicators and their associated definitions.

Inclusion/exclusion criteria and definitions

In the literature search we included only RCTs, quasi-experimental trials or before–after studies that reported interventions to improve patient-related quality and safety of care within the perioperative setting in adults >18 yr old. Interventions were defined as the implementation of best medical practices, as developed locally or available through best practice guidelines or protocols. To be considered for inclusion, studies also had to report the use of one or several clinical indicators to measure intervention effectiveness (outcome indicators).

Trials that assessed drug, device, or new procedure effectiveness or reported adverse events as secondary outcomes were excluded. We also did not consider studies that used outcomes related exclusively to intensive care or surgical care

and those reporting comfort measures (pain, postoperative nausea and vomiting, sedation, immobility) or patient-reported outcomes (quality of life, return to work, functional assessment, satisfaction), as these were reviewed by other groups of the StEP initiative. Detailed definitions used are available in [Supplementary material S2](#).

Literature search and data extraction

We performed a systematic search on MEDLINE, EMBASE, and the Cochrane Database for studies published between January 1, 2000 and March 30, 2016 in core clinical journals as defined by the National Library of Medicine.¹⁷ To identify potentially eligible studies according to title and abstract content, two authors (GH and SB) independently performed the review with the help of two professional librarians. The reference lists of retrieved articles were also searched for additional studies. If definitions of indicators in retrieved studies were poorly defined, additional searches were performed using Web of Science for abstracts of conferences, Google Scholar, professional organisations, and quality improvement initiatives websites (grey literature) in order to retrieve original definitions of those indicators. We did not apply any language restriction. A detailed description of the search strategies used is provided in [Supplementary material S3](#).

Selected articles were independently analysed by the two authors, and clinical indicators were extracted according to a standardised extraction and coding template ([Supplementary material S5](#)). Redundancy between clinical indicators was solved by aggregation into a single indicator and composite measures were excluded. The overall process was performed by two authors (GH, SB) during consensus meetings. Discrepancies were resolved by consultation between these two authors. Risk of bias in studies was not assessed, as the purpose of the review was not assessing clinical intervention effectiveness, but to identify the scope, definitions, and validity of clinical indicators currently in use. Likewise, quantitative data synthesis or meta-analysis was not performed.

Delphi process

We used a Delphi method to gain consensus around the clarity, reliability, and validity of each indicator as a measure of patient quality in perioperative care to be used in future clinical trials.^{18,19} The final list of retrieved clinical indicators and associated definitions were provided to members of the StEP Initiative Steering Committee, subgroup coordinators, and members. These were expert researchers from different anaesthesia specialties and countries (Australia, Canada, The Netherlands, UK, USA, South Africa, Senegal, Switzerland; the full list available in [Supplementary material S1](#)).

Delphi round 1

After discussion with other subgroup members, the theme subgroup chair (GH) prepared the initial list of endpoints and associated definitions retrieved from the literature according to a predefined format prepared by the StEP Steering Committee ([Supplementary material S4](#)). All members of the clinical indicators subgroup ($n=8$) and the StEP Steering Committee ($n=5$) were invited to participate.

Participants were asked to score each of the listed indicators for clinical importance using a scale of 1–9. Scores of 1–3 indicated ‘not that important or invalid’, 4–6 indicated

‘important but requires revision’, and 7–9 ‘critical for inclusion’. Participants were offered the option to select ‘not applicable/not sure’ if they were unable to form an opinion about the importance or not of the clinical indicator. Participants had 2 weeks to answer before reminder emails (up to 3) were sent to prompt completion of the survey. For each indicator, participants were also invited to add any comments, suggestions for modifications of existing definitions that they believed were important. Individual indicator scores were then calculated using mean, median, and range of scores. Comments and suggestions provided by participants were collated to be integrated to the second Delphi round.

Delphi round 2

The theme subgroup chair (GH) selected indicators that had been rated as ‘critical’ (score ≥ 7) by at least 70% of participants to prepare the first list of indicators for Delphi round 2. Indicators rated as ‘not that important or invalid’ (score ≤ 3) or as ‘important but requiring revision’ (score >3 and <7) by at least 70% of participants were also included in the second round but clearly identified as such on a second list. Members of the clinical indicators theme subgroup also discussed Delphi round 1 results and indicators definitions and selection via email.

Participants were asked to score the clinical indicators using the same questionnaire format and rating procedure as the one used during Delphi round 1. For this second round, participants were provided with the mean scores of each clinical indicator after round 1. Comments after Delphi round 1 were also added. This stage included the entire StEP Working Group ($n=54$).

Delphi round 3

The theme subgroup Chair (GH) selected for Delphi round 3 only indicators that had been rated as ‘critical’ (score ≥ 7) by at least 70% of participants during the second round. Indicators rated as ‘not that important or invalid’ (score ≤ 3) or ‘important but requiring revision’ (score >3 and <7) were not included. If responses to the second-stage Delphi process comments section suggested that modification to endpoint definitions or rating had to be made, this was discussed within the indicator’s theme subgroup via email.

For this third round, participants were provided with the short list of selected indicators and attached definitions and all comments provided after rounds 1 and 2. They were asked to score the item using a second questionnaire (pro forma available in [Supplementary material S5](#)). The questionnaire included four rating criteria per indicator:^{20,21}

1. Validity—the degree to which the indicator measures what it purports to measure
2. Reliability—the degree of stability of the indicator when measurement is repeated under identical conditions
3. Feasibility—practicability/ease of use in the clinical setting
4. Clarity of the definition—the degree to which the clinical indicator meaning can be easily understood

For each question, participants were again asked to rate each indicator on a 1 to 9 scale with scores 1–3 indicating ‘no’, 4–6 meaning ‘unsure’, 7–9 meaning ‘yes’; meanwhile, a score of 10 meant ‘not assessable’.

At the end of the third Delphi round, indicators that had a score of 7–9 (‘yes’) for each question were automatically

selected as recommended indicators for quality measurement in clinical trials. Clinical indicators rated 4–6 ('unsure') for one or several of the four rating criteria were discussed by email within the indicator's subgroup. Those that had a score of 1–3 ('no') for any of the rating criteria were considered only as optional, but not recommended.

Each Delphi round was coordinated by the Department of Anaesthesia and Perioperative Medicine at the Alfred Hospital in Melbourne, Australia. Participants' answers to the different Delphi rounds were recorded, transformed and analysed using the statistical Package for Social Sciences (SPSS®, New York). Analyses included mean, median, number, and proportion of respondents.

Results

A total of 20 058 reports were identified, of which 351 were selected for further analyses. After full content assessment and exclusion of duplicates, observational studies, costs analysis studies, surgical, new technique, or drug-related studies, 120 reports were considered. Of these, 16 further publications were excluded as no information and definition of any quality indicator was provided. The final analysis included 104 reports and 167 clinical indicators (Fig. 1). The full list of clinical indicators identified is provided in [Supplementary material S6](#).

After full assessment and consensus by the two raters (GH, SB), the final list was reduced to 101 indicators after removal of additional duplicates, composite scores, ICU, and surgical indicators and outcome measures that were assessed by other StEP-COMPAC Groups. The final list was carried forward to the Delphi process. Participation to the different Delphi rounds was 100% ($n=13$), 68% ($n=54$), and 85% ($n=6$), respectively. Results of the Delphi rounds 1, 2, and 3 are shown in [Table 1](#). None of the 101 indicators selected in the first Delphi round were removed for the second Delphi round. For the third Delphi round, all clinical indicators rated as important by more than 70% of the participants were carried forward except: 'Pulmonary embolism within 30 days of surgery', 'Myocardial infarction within 30 days of surgery', 'Myocardial infarction within 30 days of surgery (with or without in-hospital mortality)', and 'Major adverse cardiac event within 30 days of surgery' (available in [Supplementary material S7](#)). These indicators were considered by several subgroup chairs and Delphi participants as redundant with outcomes assessed by other subgroups or too poorly defined to allow further assessment.

The final list included eight clinical indicators rated for validity, reliability, usability, and clarity of definitions, and is provided in [Table 2](#). All indicators except 'Surgical Site Infection rate at 30 days after surgery according to CDC criteria' were considered by at least 50% of the experts as

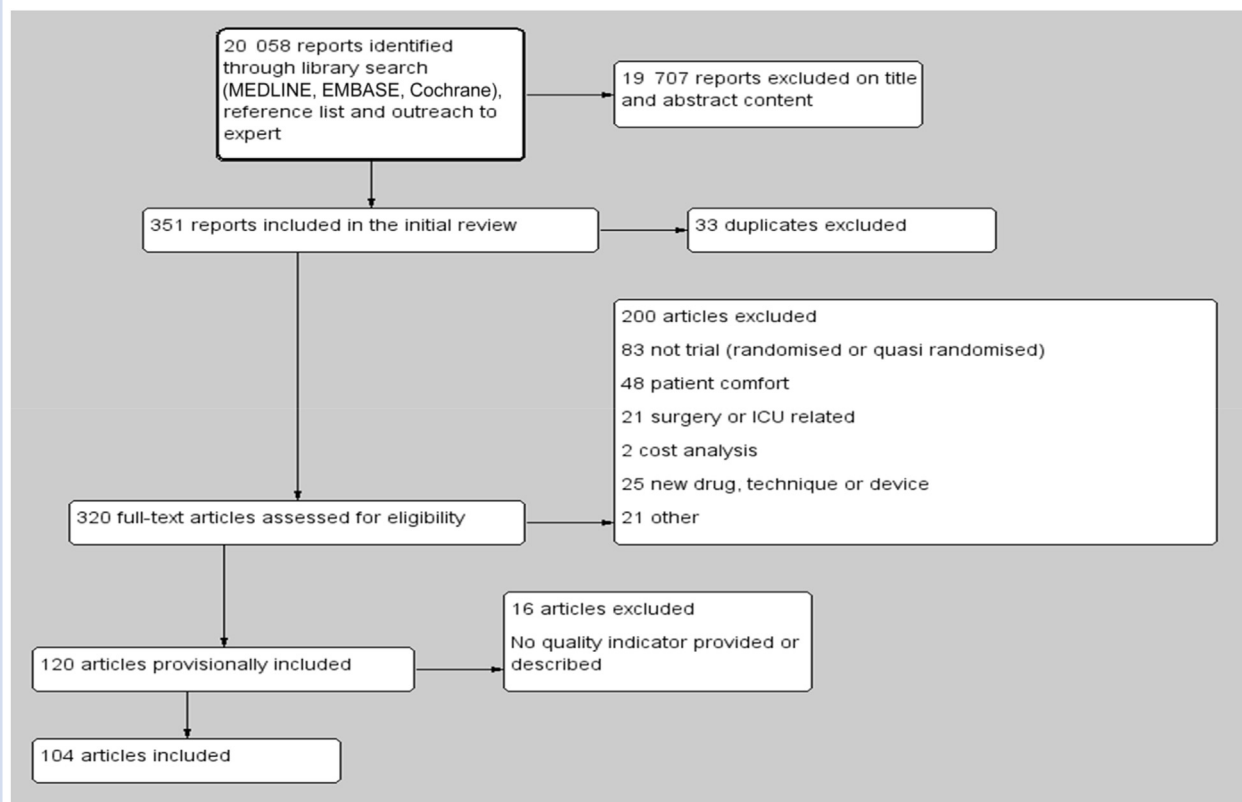


Fig 1. Flow diagram of literature search for systematic reviews of studies assessing quality improvement initiatives and using clinical indicators as study outcome.

Table 1 Results of the three Delphi rounds. ncf, not carried forward; N, total number; (%) proportion

Clinical indicator's generic name***	Delphi round 1 (N=13)			Delphi round 2 (N=54)			Delphi round 3 (N=6)		
	Unsure (n)	Median score	Score ≥ 7 (%)	Unsure (n)	Median score	Score ≥ 7 (%)	Unsure (n)	Median score	Score ≥ 7 (%)
1. Absence of falls after surgery	0	6	30	2	4	17	—	ncf	—
2a. Surgical site infection (definition 1)	1	7	70	1	7	65	—	ncf	—
2b. Surgical site infection (definition 2)	2	7	53	0	5	20	—	ncf	—
2c. Surgical site infection (definition 3)	1	6.5	46	0	5	0	—	ncf	—
2d. Surgical site infection (definition 5)	1	6.5	46	0	5	6	—	ncf	—
2e. Surgical site infection	0	8	92	0	8	89	0	8	92
3a. Wound infection within 30 (or 90) days of surgery (definition 1)	2	7	46	0	5	16	—	ncf	—
3b. Wound infection within 30 (or 90) days of surgery (definition 2)	2	7	61	0	6	38	—	ncf	—
3c. Wound infection within 30 (or 90) days of surgery (definition 3)	3	7	46	0	5.5	27	—	ncf	—
3d. Wound infection within 30 (or 90) days of surgery (definition 4)	1	7	61	0	6	34	—	ncf	—
4. Number of patients who received antibiotics beyond the recommended postoperative period	2	5	15	1	4	7	—	ncf	—
5. Quality of laryngoscopy and tracheal intubation	1	5	30	1	3	0	—	ncf	—
6a. Acute postoperative renal dysfunction (definition 1)	1	6	46	3	5.5	19	—	ncf	—
6b. Acute postoperative renal dysfunction (definition 2)	0	6	46	2	6	45	—	ncf	—
6c. Acute postoperative renal dysfunction (definition 3)	1	4.5	0	1	4	8	—	ncf	—
6d. Acute postoperative renal dysfunction (definition 4)	1	7	53	2	5.5	30	—	ncf	—
6e. Acute postoperative renal dysfunction (definition 5)	0	6	38	2	6	21	—	ncf	—
7. Acute stroke after carotid endarterectomy	0	6	46	2	6	31	—	ncf	—
8a. Stroke within 30 days of surgery (definition 1)	0	6	23	1	6	39	—	ncf	—
8b. Stroke within 30 days of surgery (definition 2)	0	7	92	1	8	90	0	7	92
9. Inadequate regional block	1	7	53	1	5	21	—	ncf	—
10. Failed regional block	0	7	61	1	5	21	—	ncf	—
11. Failed spinal block	2	4.5	15	1	4	3	—	ncf	—
12. Failed epidural	0	7	53	1	5	9	—	ncf	—
13. Unintended epidural vessel penetration	0	6	23	1	5	9	—	ncf	—
14. Incomplete epidural block	0	7	46	1	5	19	—	ncf	—
15. Very high or total spinal block as a result of accidental intrathecal injection	0	6	38	1	6	23	—	ncf	—
16a. Postdural puncture headache (definition 1)	1	6	30	1	5.5	16	—	ncf	—
16b. Postdural puncture headache (definition 2)	2	7	38	1	6	23	—	ncf	—
16c. Postdural puncture headache (definition 3)	1	7	46	1	7	53	—	ncf	—
17a. Number of intubation attempts (definition 1)	0	7	46	2	5.5	28	—	ncf	—
17b. Number of intubation attempts (definition 2)	1	5	7	2	4	3	—	ncf	—
18. Failed attempt of intubation	2	6	30	2	5	17	—	ncf	—
19. Mucosal trauma after intubation	1	4.5	15	2	5	9	—	ncf	—
20. Postoperative hoarseness	0	6	30	1	5	13	—	ncf	—
21. Recovery room airway complications	0	6	23	2	5	9	—	ncf	—
22. Intraoperative airway complications	0	6	23	2	5	13	—	ncf	—
23. Vocal cord injuries	1	6.5	38	2	6	41	—	ncf	—
24. Composite pharyngolaryngeal adverse events	0	6	38	2	6	31	—	ncf	—

Continued

Table 1 Continued

Clinical indicator's generic name***	Delphi round 1 (N=13)			Delphi round 2 (N=54)			Delphi round 3 (N=6)		
	Unsure (n)	Median score	Score ≥7 (%)	Unsure (n)	Median score	Score ≥7 (%)	Unsure (n)	Median score	Score ≥7 (%)
25. Securement of an effective airway	0	6	38	2	5	15	—	ncf	—
26. Hypoxia during mask ventilation or tracheal intubation	0	6	30	1	5	12	—	ncf	—
27. Bronchial injuries	1	4.5	7	1	4	3	—	ncf	—
28. Postoperative respiratory failure	0	7	61	1	7	69	—	ncf	—
29. Pulmonary complications	0	6	38	2	6	31	—	ncf	—
30. Pulmonary embolism within 30 days of surgery	0	7	69	1	7	73	—	ncf	—
31. Prolonged mechanical ventilation	0	7	61	0	7	59	—	ncf	—
32. Venous thromboembolism	0	7	61	1	7	69	—	ncf	—
33. Atelectasis within 14 days	0	5	15	0	4	3	—	ncf	—
34. Pneumonia within 30 days of surgery	0	7	84	0	7	67	—	ncf	—
35. Aspiration of gastric content	1	6	38	1	6	37	—	ncf	—
36. Proportion of transfused patients	0	6	38	2	7	50	—	ncf	—
37. Risk of receiving any fresh frozen plasma in-hospital	1	6	15	1	5.5	23	—	ncf	—
38a. Awareness (definition 1)	0	6	30	2	6	41	—	ncf	—
38b. Awareness (definition 2)	1	6	38	2	6	39	—	ncf	—
39a. Time to orientation (definition 1)	0	6	46	2	5	0	—	ncf	—
39b. Time to orientation (definition 2)	1	6	23	1	5	0	—	ncf	—
40. Postoperative CNS failure	3	4	23	1	5	16	—	ncf	—
41. Postoperative cardiocirculatory failure	1	4	23	2	4	16	—	ncf	—
42. New postoperative deterioration in neuropsychologic performance	1	5	30	1	5	16	—	ncf	—
43. Postoperative residual curarisation	0	6	38	2	6	26	—	ncf	—
44. Reoccurrence of neuromuscular blockade	0	6	23	3	6	23	—	ncf	—
45. Recovery of the train of four ratio	0	5	23	3	5	7	—	ncf	—
46. Perioperative hypothermia (definition 1)	0	7	53	2	7	57	—	ncf	—
47. Postoperative hypothermia (definition 2)	0	6	46	2	6	45	—	ncf	—
48. Surgeon satisfaction	0	4	30	2	4	3	—	ncf	—
49. Surgical conditions for thoracic surgery	0	6	46	2	4	6	—	ncf	—
50. Surgical space conditions	0	6	38	2	4.5	3	—	ncf	—
51. Hypoglycaemic events	0	7	69	1	7	67	—	ncf	—
52. Incidence of postoperative delirium during the postoperative hospitalisation period	0	6	46	2	7	68	—	ncf	—
53. Postoperative haematology failure	2	6	15	4	4	3	—	ncf	—
54. Postoperative hepatic failure	0	5	23	3	5	10	—	ncf	—
55. Cardiac arrest within 30 days of surgery Rating	1	7	53	1	7	59	—	ncf	—
56. Non-fatal cardiac arrest	0	7	76	1	7	65	—	ncf	—
57. Non-fatal myocardial infarction	1	7	53	2	7	63	—	ncf	—
58. New myocardial ischaemia	1	6.5	46	1	7	55	—	ncf	—
59. New clinically important atrial fibrillation	1	7	69	1	7	66	—	ncf	—
60a. Myocardial infarction within 30 days of surgery (definition 1)	1	7	69	1	7	75	—	ncf	—
60b. Myocardial infarction within 30 days of surgery (definition 2)	0	7	61	1	7	77	—	ncf	—
60c. Myocardial infarction within 30 days of surgery	2	8	46	1	6	47	—	ncf	—
61. Major adverse cardiac event within 30 days of surgery	0	7	61	1	7	77	—	ncf	—
62. Cardiovascular death within 12 months of surgery	0	6	46	2	6	31	—	ncf	—
63. 30 day mortality after coronary artery bypass grafting	1	8	84	2	8	87	1	8	84
64. Adjusted 30 day mortality after carotid endarterectomy	1	7	61	1	6	35	—	ncf	—

Continued

Table 1 Continued

Clinical indicator's generic name***	Delphi round 1 (N=13)			Delphi round 2 (N=54)			Delphi round 3 (N=6)		
	Unsure (n)	Median score	Score ≥ 7 (%)	Unsure (n)	Median score	Score ≥ 7 (%)	Unsure (n)	Median score	Score ≥ 7 (%)
65. Number of patients who died or experienced complications until hospital discharge	0	7	53	1	7	58	–	ncf	–
66a. Operative mortality (definition 1)	1	7	61	0	7	68	–	ncf	–
66b. Operative mortality (definition 2)	1	8	70	0	8	80	1	8	70
67. Incidence of 30 day overall postoperative morbidity	2	6	38	1	6	39	–	ncf	–
68. Postoperative major complications (in-hospital or within 30 days)	1	7	53	0	7	57	–	ncf	–
69. Severe hypotension	2	6	30	0	5	20	–	ncf	–
70. Severe nausea and vomiting	2	6	38	2	6	21	–	ncf	–
71. Sum of all prespecified complications	3	4.5	7	0	4	10	–	ncf	–
72. Postoperative morbidity	1	7	69	0	6	42	–	ncf	–
73. Safe to discharge from the surgical suite	1	7	69	3	6	41	–	ncf	–
74. Time to eligibility for PACU discharge	1	6	38	3	6	37	–	ncf	–
75. Time to fulfilment of criteria for leaving the operating room	1	6	38	2	6	23	–	ncf	–
76. Admission to the intensive care unit within 14 days	1	7	61	1	7	76	1	7	76
77. Eligibility to bypass the PACU, i.e. 'fast tracking'	1	6.5	46	3	6	20	–	ncf	–
78. Readmission to ICU or IMC (Immediate Care Unit)	2	7	53	0	7	69	–	ncf	–
79. Need for patient readmission	1	8	92	0	8	92	1	8	92
80a. Length of stay (definition 1)	2	7	76	0	7	75	2	7	76
80b. Length of stay (definition 2)	3	8.5	61	0	8	77	3	8.5	61

valid (score ≥ 7). The same rating (score ≥ 7) for reliability was provided for all indicators except 'Surgical Site Infection rate at 30 days after surgery according to CDC criteria' and 'stroke within 30 days of surgery'. The majority of experts ($\geq 50\%$) also rated the final list of indicators as easy to use (score ≥ 7) and the selected endpoints as having clear definitions except 'admission to the intensive care unit within 14 days'.

Discussion

After a systematic review and Delphi process to achieve consensus from a broad range of experts involved in perioperative clinical studies, we identified eight key clinical indicators to measure effectiveness of interventions (i.e. best practice protocols and guidelines) aimed at improving quality and safety.

Identification of a list of outcome indicators used as endpoints in trials through a systematic search of existing literature is an important step in the process of standardising endpoint measures for trials assessing quality and safety improvement initiatives. However for these outcome indicators to be reliably used for such purpose, they need first to have relevance for the clinical context in which they are used. We therefore asked panel experts to prioritise indicators for their use in the clinical setting and to agree on the most appropriate definitions. Secondly, as true quality and safety indicators need to reflect the care provided rather than pre-existing diseases, we asked experts to assess these indicators for their likelihood to reflect an issue in the quality of care provided. We used a methodology developed in social sciences: face validation.²¹

Face validity testing is based on field expert consensus.^{15,22} Field experts literally 'look' at the indicator and agree not only on its definition and meaning and but also validity. This is done through a consensus development process, where experts discuss meaning and definitions of clinical indicators. For this study we used the Delphi method. This entire standardised process has been developed and validated by the University of California and the American Institute of research and development.¹⁶ As a result of this iterative process, eight clinical outcome indicators were selected, their definitions clarified, and their 'face' validity confirmed for nearly all indicators. These indicators refer to harmful events likely to be caused by the care provided rather than by pre-existing disease in line with the WHO definition of patient safety 'the absence of preventable harm to a patient during the process of health care'.²³ In addition, a number of them relate to patient trajectories in hospital such as readmission to hospital, unplanned admission to ICU, and length of stay. This reinforces the likelihood of unexpected events related to care provided that may have led to prolonged stay or unexpected admission to ICU or hospital.

Most of these indicators were also considered by experts as reliable, clearly defined, and usable as endpoints in clinical trials assessing quality improvement interventions.

While a large number of clinical indicators is available in the literature or within lists provided by hospital accreditation bodies, their level of validity is often limited and definitions are often lacking or sometimes conflicting. Heterogeneity in outcome definitions can significantly impact on the reliability of clinical trials performed in the area and limit further development of scientific evidence to guide improvement initiatives.²⁴ It can also limit comparison between studies and

Table 2 Results of the final Delphi round on validity, reliability, feasibility, and clarity of definitions of selected indicators. *CDC criteria: a surgical site infection (SSI) is an infection that occurs after surgery in the part of the body where the surgery took place. Surgical site infections can sometimes be superficial infections involving the skin only. Other surgical site infections are more serious and can involve tissues under the skin, organs, or implanted material. Symptoms include: Redness and pain around the area where you had surgery; Drainage of cloudy fluid from your surgical wound; Fever. †Validity. Does the endpoint measures what it purports to measure? ‡Reliability. Is the endpoint reproducible and does it have stability when measurement is repeated under identical conditions? §Feasibility. Can the endpoint data be collected and used easily by research staff with some training, without undue effort or risk of missing data? ¶Clarity of the definition. Does the endpoint have a meaning that can be easily understood?

Clinical indicator's generic name and definition	Validity [†]			Reliability [‡]			Feasibility [¶]			Clarity of the definition [§]		
	Unsure (n)	Median score	Score ≥7 (%)	Unsure (n)	Median score	Score ≥7 (%)	Unsure (n)	Median score	Score ≥7 (%)	Unsure (n)	Median score	Score ≥7 (%)
Surgical site infection	4	5.5	16	4	5	16	2	6	50	2	7	66
<i>Surgical site infection rate at 30 days after surgery according to CDC criteria*</i>												
Stroke within 30 days of surgery	2	6	50	3	5.5	33	2	7	66	1	7.5	83
<i>Cerebral haemorrhage on CT or MRI, or new neurological signs (paralysis, weakness, or speech difficulties) lasting 24 h or leading to earlier death</i>												
30 day mortality after coronary artery bypass grafting	3	6.5	50	1	8	83	0	8	100	0	8.5	100
<i>Death within 30 days of coronary artery bypass grafting</i>												
Operative mortality	2	7	66	1	7.5	83	0	8	100	0	8	100
<i>Death within 30 days of surgery</i>												
Admission to the intensive care unit within 14 days	3	6.5	50	3	6.5	50	3	6.5	50	4	6	33
<i>Admission to ICU within 14 days of surgery and not part of the postoperative care</i>												
Need for patient readmission	0	7.5	100	1	7.5	83	0	7.5	100	1	7.5	83
<i>Readmission to hospital within 30 days of surgery</i>												
Length of stay (LOS)	3	6.5	50	3	6	50	0	7	100	1	7	83
<i>LOS defined as the postoperative hospital LOS and calculated by subtracting using date of hospital discharge to date of surgery</i>												
Length of stay	2	7	66	2	6	50	1	7	83	0	8	100
<i>LOS is defined as the number of days from the day of surgery to hospital discharge or death</i>												

further combination of studies in meta-analyses.²⁵ We tried to identify a usable list of relevant clinical outcome indicators that are clearly defined, reliable, and valid as quality and safety measurement tools to be used in clinical trials and health services research addressing patient quality and safety issues. It should be emphasised that the feasibility of measuring a specific endpoint in any trial will also depend on other factors that were not specifically measured in this study and that could still hinder use of the recommended indicators. These include especially the complexity of the clinical trial and the available budget. Although many endpoints and indicators can be recommended, it is ultimately the workload to collect them and the available budget to do so that will decide.

Several limitations of this study have to be mentioned. First is the methods used. Because it is based on a systematic review, it only enables identification and rating of existing indicators. We did not develop a list of new indicators for future use as endpoints in perioperative trials. However, an advantage of the chosen method is that indicators for which consensus definitions are achieved will be immediately available, enabling researchers and clinicians to derive more value and applicability from findings of published perioperative research.

Second is that clinical indicators often refer to postoperative complications. They can consequently easily be confused with broader clinical trials outcome measures such as postoperative respiratory failure or acute postoperative myocardial infarct. This may lead to some confusion as to the specificity of selected indicators for quality and safety measurement. To limit this risk, we used strict inclusion criteria and selected only reports of interventions that aimed specifically at improving patient-related quality and safety of care within the perioperative setting. To identify these indicators we used search terms that directly related to quality and safety such as 'iatrogenic complications', 'adverse drug reaction', and 'complication avoidance'. We also asked panel experts whether 'faced with the indicator' they could conclude that it was a valid, reliable, usable, and clearly defined measure of patient quality and safety in perioperative medicine.

Third is that the purpose of the StEP international initiative was to identify and provide straightforward, clinically sensible, and valid consensus definitions for a comprehensive set of trial endpoints.²⁶ As a result, there was some overlap between the different outcomes identified by the different subgroups participating in the initiative. This was the case in our study for outcomes related to postoperative complications (e.g. aspiration), that could be interpreted as conventional outcome measures whereas they can also relate to quality and safety. To solve this issue, discussions took place during the Delphi process between subgroup chairs to reach a consensus between groups as to which group should assess overlapping endpoints.

Fourth is that for 'Surgical Site Infection rate at 30 days after surgery according to CDC criteria' both validity and reliability were limited. Only 16% of the experts provided a score ≥ 7 for validity and reliability. This was not because experts found that association with quality was unclear. Nosocomial infection is a valid and worldwide used clinical indicator. However, they found that the definition of an infection through 'redness and pain with cloudy fluid secretion' was not specific and the risk of misdiagnosis was high. As a result we recommend its use as an endpoint for quality and safety measurement with some caution.

Fifth is that our literature search was limited to core clinical journals. Although this approach increased the likelihood of

retrieving the most clinically relevant and read publications, it limited at the same time the scope of the literature search process.

Finally, as for any diagnostic tool in medicine, specificity of measurement tools (here clinical indicators) rarely reaches 100%. In our study, only 'readmission to hospital within 30 days of surgery' was considered by 100% of the experts questioned as having very high validity. For all other endpoints, the highest level of validity was attributed by experts in 16%–66% of the cases, depending on the indicator assessed. This suggests that these indicators may also in some circumstances measure other dimensions of care (i.e. patient decision to leave hospital earlier than recommended; severe pre-existing disease leading to early postoperative mortality). However, in most other cases, the outcome indicators identified can be confidently used as endpoints in clinical trials and health services research assessing patient quality and safety improvement initiatives.

Despite these limitations, we identified a number of standardised endpoints for quality and safety measurement to be used in future studies assessing effectiveness of interventions aimed at improving medical practices through best practice guidelines or protocols. This study should improve both consistency in the use of perioperative clinical indicators and reliability of clinical trial results. This should translate into improved interpretation of study results and better translation into clinical practise.

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 Wijesundera (Canada).

Authors' contributions

Study concept: GH, SB, PM.

Protocol development: GH, SB, TC, DS, MPWG, MN, LF, PM.

Writing of the initial draft of the manuscript: GH, SB.

Revision of manuscript: GH, SB, TC, DS, MPWG, MN, LF, PM.

Systematic review: GH, DS.

Assistance with systematic review: MB (Mafalda Burri).

Approval of the final manuscript for publication: GH, SB, TC, DS, MPWG, MN, LF, PM. All authors participated in the Delphi survey and analysis of the responses, critical review and revisions of the protocol and manuscript.

Acknowledgements

We thank Mafalda Burri, librarian at the University of Geneva, for assistance in the literature search, Angela Lowe and Sophie Wallace (Alfred Hospital, Melbourne) for assistance with the Delphi surveys, and all members of the StEP-COMPAC Group who responded in a timely fashion. We thank the *British Journal of Anaesthesia* for sponsoring a workshop on this topic at the Monash University Prato Centre, Italy, in June 2015.

Declarations of interest

PSM is an editor and TC and GH are members of the associate editorial board of the *British Journal of Anaesthesia*. None of the

other authors report any competing interests relating to the topic of this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.bja.2019.04.041>.

References

1. Genovitch-Richards J. Quality management organizational structures: history and trends. *J Healthc Qual* 1994; **16**(22–5): 8–9
2. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003; **15**: 523–30
3. Chazapis M, Gilhooly D, Smith AF, et al. Perioperative structure and process quality and safety indicators: a systematic review. *Br J Anaesth* 2018; **120**: 51–66
4. Joint Commission of Healthcare Organizations, National library of healthcare indicators. *Health plan and network edition*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organizations; 1997
5. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; **260**: 1743–8
6. NELA Project Team. Second Patient Report of the National Emergency Laparotomy Audit, Available from: www.nela.org.uk/reports; 2016
7. Broder MS, Payne-Simon L, Brook RH. Measures of surgical quality: what will patients know by 2005? *J Eval Clin Pract* 2005; **11**: 209–17
8. Haller G, Myles P. Use and misuse of outcome data. *Lancet* 2004; **363**: 1912
9. Copnell B, Hagger V, Wilson SG, Evans SM, Sprivulis PC, Cameron PA. Measuring the quality of hospital care: an inventory of indicators. *Intern Med J* 2009; **39**: 352–60
10. Anesthesia Quality Institute Outcome definitions. Available at www.aqihq.org/AQI_Outcomes_Definitions_2015.
11. Australian Council on Healthcare Standards. Clinical indicators definitions. Available from: at www.achs.org.au.
12. Haller G, Stoelwinder J, Myles PS, McNeil J. Quality and safety indicators in anesthesia: a systematic review. *Anesthesiology* 2009; **110**: 1158–75
13. Benn J, Arnold G, Wei I, Riley C, Aleva F. Using quality indicators in anaesthesia: feeding back data to improve care. *Br J Anaesth* 2012; **109**: 80–91
14. Grocott MP. Standardisation of perioperative outcome measures. *Eur J Anaesthesiol* 2015; **32**: 83–4
15. Campbell SM, Braspenning J, Hutchinson A, Marshall M. Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002; **11**: 358–64
16. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *Int J Qual Health Care* 2003; **15**: i5–11
17. National Institute of Medicine. Abridged Index Medicus (AIM or “Core Clinical”) Journal Titles. Available from: <https://www.nlm.nih.gov/bsd/aim.html>.
18. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol* 2003; **56**: 1150–6
19. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; **311**: 376–80
20. Last J. A dictionary of epidemiology. In: *A handbook sponsored by the international epidemiology association*. 4th edn. New York: Oxford University Press; 2001
21. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th edn. Oxford; New York: Oxford University Press; 2008. p. 150–62
22. McGlynn EA, Asch SM. Developing a clinical performance measure. *Am J Prev Med* 1998; **14**: 14–21
23. World Health Organization. Patient Safety Initiative. Available from: <http://www.euro.who.int/en/health-topics/Health-systems/patient-safety>.
24. Koroshetz W. A core set of trial outcomes for every medical discipline? *Br Med J* 2015; **350**: h85
25. Ioannidis JP. How to make more published research true. *PLoS Med* 2014; **11**: e1001747
26. Boney O, Moonesinghe SR, Myles PS, et al. Standardizing endpoints in perioperative research. *Can J Anaesth* 2016; **63**: 159–68

Handling editor: H.C. Hemmings Jr