

BMJ Open Clinical indicators for reporting the effectiveness of patient quality and safety-related interventions: a protocol of a systematic review and Delphi consensus process as part of the international Standardised Endpoints for Perioperative Medicine initiative (StEP)

Sohail Bampoe,¹ Tim Cook,² Lee Fleisher,³ Michael P W Grocott,⁴ Mark Neuman,⁵ David Story,⁶ Paul Myles,⁷ Guy Haller⁸

To cite: Bampoe S, Cook T, Fleisher L, *et al.* Clinical indicators for reporting the effectiveness of patient quality and safety-related interventions: a protocol of a systematic review and Delphi consensus process as part of the international Standardised Endpoints for Perioperative Medicine initiative (StEP). *BMJ Open* 2018;**8**:e023427. doi:10.1136/bmjopen-2018-023427

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-023427>).

Received 8 April 2018

Revised 11 September 2018

Accepted 25 September 2018



© Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Sohail Bampoe;
s.bampoe@gmail.com

ABSTRACT

Introduction Clinical indicators are used to measure and quantify the safety and quality of patient care. They are also often used as endpoints in clinical trials. Definitions of clinical indicators in common use are extremely heterogeneous, limiting their applicability. As part of the international Standardised Endpoints in Perioperative Medicine initiative, this study will identify clinical indicators by systematically reviewing the anaesthesia and perioperative medicine literature, and will provide consensus, clinically useful definitions for those indicators using a Delphi process.

Methods and analysis An electronic database search will be conducted of Medline (PubMed/OVID), EMBASE and the Cochrane Library in order to meet this review's objectives that are: (1) To identify clinical indicators and their definitions used in randomised controlled trials that assess patient-related quality and safety interventions in perioperative medicine; (2) To select a shortlist of recommended indicators and definitions that are the most suitable for evaluation of quality and safety interventions following an expert-based consensus-gaining process (Delphi method) and (3) To provide a classification scale for each indicator related to its clarity of definition, validity (strength), reliability, feasibility (ease of use) and frequency of use. This systematic review protocol is reported in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols guidance.

Ethics and dissemination Ethical approval is not required for this systematic review and Delphi process. The results of this study will be disseminated to the anaesthesia and perioperative medicine clinical and academic community through national and international presentations and through publication in a peer reviewed journal.

PROSPERO registration number CRD42016042102.

Strengths and limitations of this study

- A strength of this study is its broad inclusion criteria that will allow the identification and subsequent standardisation of a large number of clinical indicators to be used as endpoints in clinical trials, thereby increasing the value and clinical applicability of perioperative research as part of the international Standardised Endpoints in Perioperative Medicine initiative.
- A limitation of this study is that systematic review methodology only allows the identification and subsequent standardisation of existing clinical indicators rather than the development of novel indicators for future use.
- A further limitation is that the literature search is limited to core clinical journals in order to retrieve the most clinically relevant and read publications.
- Another potential limitation of this study is the exclusion of studies reporting clinical indicators before the year 2000.
- A strength of this study is the broad range of cross-specialty clinicians participating in the Delphi process.

INTRODUCTION

The drive to improve the quality of care provided in health systems around the world¹ has led to increased scrutiny in the measurement of patient outcomes. In order to inform local, national and international improvement initiatives, measurement and quantification of quality and safety of patient care are necessary.² Clinical safety and quality indicators provide clinicians, healthcare regulators and health funders with the means to

both quantitatively and qualitatively describe healthcare performance.³ They can, however, also be used in clinical trials in order to provide results that inform clinicians of best practice methods. For such purposes, it is important for clinical indicators to be consistent and reliable.

Developed initially in the manufacturing industry, indicators were first used in healthcare over three decades ago and have increasingly become used to monitor hospital organisational performance or patient care.⁴ Indicators can be classified and are routinely described as relating to healthcare structure, process or outcome.⁵ Structure indicators measure organisational resource utilisation and resilience, and are used to measure the quality of the setting in which care is delivered.⁶ Structure indicators can be used to measure an institution's capacity to respond to the healthcare requirements of the population it serves. Examples may include funding allocations, staffing levels and access to equipment and facilities such as '24 hours access to a fully staffed emergency theatre'.

Process indicators allow us to measure the quality of delivery of care. They allow for the comparison of existing practice against standards of evidence based, best practice and are most commonly used to drive improvement initiatives.⁶ Examples may include the timeliness of radiology reporting or seniority of operating surgeon. Outcome indicators are perhaps most familiar to clinicians and allow the measurement of the effects of health interventions on patients' health and well-being.³ Traditionally, outcome indicators such as mortality and morbidity have been important to clinicians and patients. However, these outcome indicators are often poor discriminators of quality of care because of the broad range of potential confounders that may affect patient outcome.³ A patient may experience poor outcomes despite exemplary care, or conversely, have excellent outcomes following low-quality care. As such, a much wider range of outcome indicators have been developed and are in use that describe a much broader set of clinical safety and quality outcomes.

Anaesthesia is a specialty that has safety at its heart. Perioperative medicine has evolved from anaesthesia out of a desire to improve outcomes for patients undergoing both elective and emergency surgery and perioperative research often uses quality and safety indicators as primary endpoints. A range of national quality improvement programmes have been developed within anaesthesia and perioperative medicine that use nationally measured structure, process and outcome indicators to drive local, institutional improvement. Examples include the National Emergency Laparotomy Audit⁷ and the Perioperative Quality Improvement Programme⁸ in the UK and the American College of Surgeons National Surgical Quality Improvement Programme⁹ in the USA.

A search of the literature will reveal a plethora of clinical indicators used in healthcare that have been developed over the last three decades.¹⁰ However, for a quality indicator to be useful, it must demonstrably lead to improved quality of care. There is considerable variation in the methods used to develop clinical indicators,^{3 11} resulting

in lack of consistency in the key attributes required of the ideal indicator, validity, reliability, applicability and relevance.¹²

In their 2017 systematic review, Chazapis *et al* identified 261 structure and process indicators in perioperative care.¹³ Notably, they found that the majority of indicators in use had none or very little supporting evidence beyond face validity (expert opinion). Haller *et al* identified 108 clinical indicators available for quality and safety measurement in anaesthesia care alone.¹⁴ Patient safety (83%) and effectiveness (68%) were the two dimensions most often addressed by these clinical indicators. However, only 40% of these clinical indicators also had a level of validity beyond face validity and many of them had inconsistent definitions across publications or in different national programmes. For instance, according to the Australian Council on Healthcare Standards, an unplanned admission to the intensive care unit was defined as 'an unplanned admission to the intensive care unit within 24 hours of a procedure with an anaesthetist in attendance' whereas the Anaesthesia Quality Institute (USA) defined it as 'an unplanned admission to the intensive care unit within 48 hours of induction'.^{15 16}

As a result of this lack of consensus in defining clinical indicators across systems and countries, it becomes difficult to assess clinical performance and benchmark quality across hospitals. When considering the use of clinical indicators to report effectiveness of patient quality and safety improvement initiatives in health services research, an additional issue arises: the limited academic interest in clinical indicators. Consequently, these measurement tools are not always clearly identified and defined.¹⁷ Such heterogeneity and lack of precise definitions limit the applicability of perioperative research by hindering our ability to compare, contrast and combine study data in order to identify the best evidence-based practice.¹⁸

Therefore, there exists a pressing need to define with greater precision and consistency clinical indicators that can be used as endpoints in perioperative clinical trials, particularly those that assess quality and safety improvement initiatives.¹⁸

The 'Standardised Endpoints for Perioperative Medicine' (StEP) initiative is an international, multidisciplinary collaborative dedicated to the identification and standardisation of a broad range of endpoints used for research and quality improvement perioperative medicine.¹⁹ It aims to develop precisely defined, clinically valid and internationally applicable perioperative endpoints. A number of endpoint-specific subgroups of subject matter experts have been convened to use Delphi methods to explicitly define each endpoint. The 'clinical indicators' subgroup members are tasked with developing a set of consensus-based definitions for clinical indicators commonly used in the perioperative literature. It has the following objectives:

1. To identify clinical indicators and their definitions used in randomised controlled trials that assess

patient-related quality and safety interventions in perioperative medicine.

2. To select, following an expert-based consensus-gaining process (Delphi method), a shortlist of recommended indicators and definitions that are the most suitable for the evaluation of quality and safety interventions.
3. To provide a classification scale for each indicator related to its clarity of definition, validity (strength), reliability, feasibility (ease of use) and frequency of use.

By standardising the evidence base for assessing quality improvement initiatives, there may be more uniform adoption of these definitions as clinical indicators by national governments and payers in different countries.

METHODS AND ANALYSIS

The standard method developed by the University of California and the American Institute of research and development will be used.^{20 21} A systematic review of the literature will be followed by the consensus opinion of a committee of experts. This approach has been adapted to facilitate the development and validation of clinical indicators and is divided into three different steps.

STEP 1: SYSTEMATIC REVIEW OF THE LITERATURE

The first step will be to identify potential indicators. This will be done by an extensive and systematic review of the existing literature on perioperative care using strict definitions (box 1).

Box 1 Definitions used in the study

Perioperative care

All aspects of anaesthesia and perioperative care in adults other than the surgery itself. Specifically, obstetrics, pain and critical care can be included if they overlap with anaesthesia care but not if there is no perioperative or anaesthetic element.

Quality improvement intervention

Any intervention implemented at the level of a practice, hospital, health system or region, intended specifically to improve either the quality or the safety of care. The intervention includes the implementation of best medical practices, as defined locally or available through the best practice guidelines or protocols.

Clinical indicator

An explicit measure (defined by the developer) of some aspect of patient clinical care used to judge a particular clinical situation and measure the quality of the care delivered.⁶

Outcome indicators

Patient-related end results of anaesthesia and perioperative care according to Donabedian's framework and definition of quality of care.⁵

Quality of care

The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with the current professional knowledge.³⁰

Eligibility criteria

Participants/population

Only randomised controlled trials, quasi-experimental trials or before–after studies that report interventions to improve patient-related quality and safety of care within the perioperative context in adults >18 years old will be included. Furthermore, only studies published between 1 January 2000 and 30 March 2016 in core clinical journals as defined by the National Library of Medicine²² will be included.

Interventions and comparators

Interventions will include those implemented at the level of a practice, hospital, health system or region, intended specifically to improve either the quality or the safety of care. Interventions will include the implementation of best medical practices, as defined locally or available through best practice guidelines or protocols.

Outcomes

To be considered, the study will also need to use one or several clinical indicators as a measure of intervention effectiveness (outcome indicators). To be considered as a clinical indicator, the outcome measured needs to be an explicit measure of some aspect of patient clinical care (ie, safety, effectiveness) and will be used to judge a particular clinical situation and measure the quality of the care delivered by a healthcare provider.

Setting

We will consider studies from all settings, including high-income and low-income and middle-income countries.

Patient and public involvement

This study is non-clinical in nature, and therefore, patient and public involvement was not sought in the study design.

Exclusion criteria

We will exclude trials that assess drug, device or new procedure effectiveness and those that report adverse events as secondary outcomes. Studies that use patient comfort measures (pain, Post operative nausea and vomiting, sedation, immobility), patient-reported outcome measures (quality of life, return to work, functional assessment, satisfaction) and cost-related measures as clinical indicators will also be excluded. Studies that use indicators related exclusively to intensive care or surgical care will be excluded. These measures will be reviewed by other StEP subgroups.

Information sources

The initial literature search will be performed to retrieve all clinical indicators exclusively from clinical trials. Database searches will be performed of Medline (PubMed/OVID), EMBASE and the Cochrane Library. The reference lists of retrieved articles will also be searched. If definitions of indicators in retrieved studies are incompletely

defined, additional searches will be performed using web of science for abstracts of conferences, Google Scholar, professional organisations and quality improvement initiatives websites (grey literature) to retrieve original definitions of those indicators.

Search strategy

The literature search will be performed separately by two reviewers and specialised librarians. Results will be compared and integrated during a consensus meeting.

Search criteria will include a predefined list of clinical indicators extracted from systematic reviews and consensus statement papers on clinical indicators in perioperative care (box 1) and will be combined with various search terms related to quality and safety improvement initiatives, for example, [unplanned ICU admission [All Fields] AND/OR ("patient safety"[MeSH Terms] OR 'quality of care'[All Fields] OR 'quality improvement'[All Fields])]. Limits will then be applied to restrict retrieved articles to 'clinical trials' in 'core clinical journals'.

The search will cover the period between January 2000 and March 2016. A complementary handsearch of the reference lists of retrieved articles will be performed for additional citations. We will also use observational studies and grey literature to complete definitions not available in the literature search. We will apply no language restriction. Authors of original articles will be contacted if study definitions of clinical indicators used are unclear or relevant data are missing. Detailed search strategy is available in online supplementary appendices 1 and 2.

STEP 2: STUDY SELECTION AND DATA EXTRACTION

The second phase will include the selection of articles and indicator programmes for data abstraction and quality rating.

Study selection, data management and data extraction process

Two authors will independently screen titles, abstract contents and full article contents to exclude citations unrelated to the study topic, using other types of outcome measurement, being non-randomised trials or published in non-core clinical journals. For the selection process, Microsoft Endnote will be used and definitions provided in box 1 will be used.

Data extraction

Following study selection, a full analysis of the selected articles will be performed independently by two authors. These will be analysed and data will be extracted

according to a standardised extraction and coding template (table 1) with predefined categories and will include seven different categories:

1. Indicator's abbreviated/standard name.
2. Author's definition.
3. Indicator's clarity (definition provided) 0=no 1=yes.
4. Validity as a measure of quality and safety (0=no validity 1=intermediate validity 3=excellent validity)
5. Study references.
6. Other indicators not defined or secondary outcomes in the study.
7. Additional details.

Redundant clinical indicators will be aggregated and composite measures excluded. The overall process will be performed by two assessors during a consensus meeting. Both are medical doctors with health services research training and/or anaesthesia specialty training. Any disagreement between the two reviewers will be resolved by discussion until a consensus between the two assessors has been reached.

Risk of bias will not be assessed as the purpose of this review is not to assess the effectiveness of any clinical information, but instead to identify the scope, definitions and validity of clinical indicators currently in use. Similarly, quantitative data synthesis or meta-analysis will not be performed.

STEP 3: DELPHI CONSENSUS PROCESS

The purpose of the third step is to gain a consensus around the validity of each indicator as a measure of patient quality in perioperative care («face validity») and consensus definitions to be used later in clinical trials.

During the third phase, the final shortlist of retrieved clinical indicators and associated definitions will be provided to members of the StEP initiative steering committee, subgroup coordinators and members. This represents a group of experts from surgery, intensive care and different anaesthesia specialties across different countries (Australia, Canada, The Netherlands, UK, USA, South Africa, Senegal and Switzerland). We will ask for their opinion regarding the indicators' validity as measures of quality, using the Delphi method.^{23 24} This method aims to establish a consensus around an issue or a definition using successive questionnaires.

The Delphi method has the advantage of allowing each individual expert to provide their own individual opinion and thus avoid the common issues (ie, Hawthorne effect; influence of opinion leaders) associated with traditional consensus conferences.

Table 1 Data extraction framework

Indicator's abbreviated/ standard name	Developer's definition/criteria	Clarity (definition provided) 0=no 1=yes	Validity as a measure of quality and safety (0=no validity 1=intermediate validity 3=excellent validity)	References	Other indicators not defined or secondary outcomes	Additional details
---	------------------------------------	---	--	------------	--	-----------------------

Delphi round 1

Following discussion with subgroup members, the theme subgroup chair will prepare the initial list of endpoints and associated definitions retrieved from the literature according to a predefined format prepared by the StEP steering committee (table 1). This initial list will be cross-checked for consistency and usability by another member of the theme subgroup (SB). The list will then be sent to all members of the clinical indicators' subgroup and the StEP steering committee in a process facilitated by the National Institute of Academic Anaesthesia (NIAA) Health Services Research Centre (HSRC).

Participants will be asked to score each of the indicators listed for clinical importance and clarity of definition using a scale of 1–9. If they consider the indicator and associated definition to not be useful or important, they will assign the labels 1–3 'not that important or invalid'. If participants consider the indicator to be useful or important but requiring improved definition and specification, they will assign the labels 4–6 'important but requires revision'. In contrast, if the indicator is seen as crucial and well defined, participants will assign the label between 7 and 9 'critical for inclusion'. Participants will be given the option to select 'not applicable/not sure' if they are unable to offer an opinion as to whether the item is important or not.

Participants will be allowed 2 weeks for task completion. A reminder email will be sent to prompt completion of the survey. The final number of respondents will be recorded (and later reported).

For each indicator, participants will be also invited to add any other endpoints, definitions or modifications to existing definitions that they believe to be important when conducting clinical trials that use quality and safety indicators. This will be possible by adding a comment in the 'comments or suggestions for this endpoint or choice of scale' section of the initial list (table 1). All answers will be sent back to the HSRC (via website or email—instructions to follow). The HSRC will be responsible for scoring the initial questionnaire using mean, median and range of scores and also collate all comments and suggestions provided by participants.

Delphi round 2

The theme subgroup chair will systematically select indicators that have been rated as 'critical' (score ≥ 7) by at least 70% of participants (score 70% percentile ≥ 7) for the second round Delphi process. In contrast, indicators rated as 'not that important or invalid' (score ≤ 3) by at least 70% of participants (score 70% percentile ≤ 3) or as 'important but requiring revision' (score 70% percentile between >3 and <7) will be included for the second round only if suggested by any of the member of subgroup or the StEP working group. Additional endpoints or suggested modified definitions will also be discussed at this stage of the Delphi process.

The list of clinical indicators will be carried forward to the Delphi second round. This stage will include the entire StEP working group. This will be done with the help of the HSRC.

For this second round, participants will be provided with the number of respondents and scores for each selected item following round 1 in the 'critical' (score ≥ 7) category. They will be asked to consider the responses from other Delphi participants and to rescore the item using the same original questionnaire (table 2). A second list of indicators rated as 'not that important or invalid' (score ≤ 3) or as 'important but requiring revision' (score 70% percentile between >3 and <7) will be provided to the entire StEP working group including suggestions made by email and participants invited to rate these indicators only if considered as critical by any group member. Responses will be collected and sent to the subgroup chair by the HSRC.

Delphi round 3

The theme subgroup chair will only select indicators that have been rated as 'critical' (score ≥ 7) by at least 70% of participants (score 70% percentile ≥ 7) for the third round Delphi stage. Indicators rated as 'not that important or invalid' (score ≤ 3) or 'important but requiring revision' (score between >3 and <7) will not be included unless at this stage any participant suggests it should still be included in the third round. If responses to the second stage Delphi process suggest that modification to endpoint definitions or rating has to be made, this will be resolved within the indicator's theme subgroup via email discussion for the third stage of the Delphi process.

Table 2 Delphi round 1 and 2 format: example of absence of falls following surgery

Endpoint and definition	Not that important or invalid			Important but requires revision			Critical for inclusion			NA or not sure	Score
	1	2	3	4	5	6	7	8	9		
1. Absence of falls following surgery Definition: absence of falls (any documented fall during hospital stay)											
Rating											

NA, not applicable.

Table 3 Absence of falls following surgery

Questions	No			Unsure			Yes			NA
	1	2	3	4	5	6	7	8	9	10
Is this indicator valid?										
Is this indicator reliable?										
Is this indicator easy to use?										
Is this indicator definition easy to understand?										
Comment										
Definition: absence of falls (any documented fall during hospital stay).										
NA, not applicable.										

The questionnaire for the third stage of the Delphi process will be sent with the assistance of the NIAA HSRC to all members of the clinical indicators' subgroup and the StEP steering committee. For this third round, participants will be provided with the number of respondents and scores for each selected item following round 2. They will be asked to consider the responses from other Delphi participants and score the item using a second questionnaire (table 3).

The questionnaire includes four rating criteria per indicator^{25 26}:

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 context.
4. Clarity of the definition—The degree to which the clinical indicator meaning can be easily understood.

For each question, participants will be asked to rate the indicator on a 1–9 scale with scores between 1 and 3 meaning 'no', 4–6 meaning 'unsure', 7–9 meaning 'yes'. Score 10 is not assessable. Details are provided in table 3.

At the end of the third Delphi round, indicators that have a score between 7 and 9 ('yes') for each question will be automatically selected as recommended indicators for quality measurement in clinical trials. Those that have been rated as 4–6 ('unsure') for one or several of the four rating criteria will be discussed by email within the indicator's subgroup. Those that have a score between 1 and 3 ('no') for any of the rating criteria will not be recommended, but still provided as a possible option for researchers. If responses to this final stage suggest that modification to endpoint definitions or rerating has to be conducted, this will also be resolved within the indicator's theme subgroup via email discussion.

Expected results

The systematic review will provide a hierarchical long-list of candidate endpoints with an associated rating for each clinical indicator. This will be followed by a shortlist of selected indicators and consensus definitions according to expert opinion. It will also provide researchers with a validity, reliability, feasibility and clarity rating and recommendations of use for indicators in quality improvement trials.

Ethics and dissemination

Results of this StEP consensus-forming process to provide recommended indicators and definitions will be presented at national and international conferences, as well as published in peer-reviewed journals. It will provide both researchers and important healthcare system stakeholders with valid and consensus indicators and definitions to be used both in clinical trials and in quality improvement initiatives.

DISCUSSION

This paper describes the protocol of a systematic review and associated Delphi process designed to identify, select and assess the most valid and suitable clinical indicators to be used for the assessment of quality and safety interventions.

There is currently a large number of clinical indicators with various and sometimes conflicting definitions that hinder their use for the assessment of perioperative care and more extensively, health services research. This heterogeneity also has an impact on the generation of scientific evidence that can guide clinical practice. In particular, clinical trials in the perioperative area can sometimes provide different conclusions owing to significant variability in outcome definitions.²⁷ This may reduce the applicability of this research to everyday practice. This systematic review and consensus-gathering process is part of the international StEP collaboration and is an important step in the process of identifying and standardising clinical indicators to be used as health service research and quality improvement endpoints.

A limitation of this study is that the current methodology does not allow for the definition or development of novel indicators for future use as endpoints in perioperative trials. Because it is based on a systematic review, the current method allows only for the appraisal and standardisation of existing indicators. An alternative method could be to first identify broad areas of academic and clinical interest and then embark on an attempt to define indicators in current use in these areas, and indicators for future use. However, an advantage of the chosen method for our study 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.

Another limitation is that clinical indicators can be sometimes confused with the broader outcome measures

often used in clinical trials, especially when they relate to postoperative complications (ie, respiratory failure, acute postoperative myocardial infarct). This may lead to an extensive literature search with a large number of outcomes retrieved that lack the specificity for quality and safety measurement. This is why we will use broad, but strict inclusion criteria and select only randomised controlled trials, quasi-experimental trials or before–after studies that report interventions to improve patient-related quality and safety of care within the perioperative context.

Another limitation is that the purpose of the StEP international initiative is to identify and provide straightforward, clinically sensible and valid consensus definitions for a comprehensive set of trial endpoints^{28 29} and as a result, there is a risk of overlap between the different outcomes identified by the different subgroups participating in the StEP international initiative. This is particularly true for outcomes related to postoperative complications (eg, aspiration), where they may be interpreted as conventional research outcome measures, but they also relate to quality and safety. This may bring some confusion as to which subgroup should deal with some outcomes (eg, the quality and safety indicator subgroup or the respiratory outcome subgroup). This may also add complexity to the consensus-gathering process as to which group should provide the definitive consensus definition. This will be resolved by discussion between subgroup chairs to decide which group should deal with overlapping measures.

Finally, the literature search is limited to core clinical journals. While this approach increases the likelihood of retrieving the most clinically relevant and read publications, it also limits the scope of the literature search process.

Once complete, the study will be published in a peer-reviewed journal. The aim of this study and the StEP initiative is to disseminate these standardised endpoints to the anaesthesia and perioperative medicine clinical and academic community to use in clinical trials and to inform clinical practice. This will be achieved through presentation at national and international conferences, and through publication in peer-reviewed journals.

CONCLUSION

Identifying standardised endpoints for quality and safety measurement should improve the consistency of perioperative clinical indicators used in research and quality improvement, and also therefore the reliability of their results. This should translate into improved interpretation of study results and better translation into clinical practice.

Author affiliations

¹Centre for Perioperative Medicine, University College London, London, UK

²Anaesthesia and Intensive Care, Royal United Hospitals Bath NHS Foundation Trust, Bath, UK

³Anesthesiology and Critical Care, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, USA

⁴Anaesthesia and Intensive Care, University of Southampton, Southampton, UK

⁵Anesthesiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

⁶Perioperative and Pain Medicine Unit, Melbourne Medical School, The University of Melbourne, Parkville, Victoria, Australia

⁷Anaesthesia and Perioperative Medicine, Monash University, Melbourne, Victoria, Australia

⁸Anaesthesia and Intensive Care/Epidemiology, Geneva University Hospital, Geneva, Switzerland

Contributors SB, GH, TC, DS, MPWG, MN, LF and PM are members of the StEP Clinical Indicators working group. GH, SB, TC, DS, MPWG, MN, LF and PM conceived and designed the study. GH and SB wrote the initial draft of the protocol. TC, DS, MPWG, MN, LF and PM revised the protocol for intellectual content. SB, GH, TC, DS, MPWG, MN, LF and PM approved the final manuscript for publication.

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

Competing interests None declared.

Patient consent Not required.

Ethics approval As the review is based on previously published literature, it does not require formal institutional ethics approval.

Provenance and peer review Not commissioned; externally peer reviewed.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Institute of Medicine. *Crossing the Quality Chasm. A New Health System for the 21st Century*. Washington: National Academy Press, 2001.
2. Scobie S, Thomson R, McNeil JJ, *et al*. Measurement of the safety and quality of health care. *Med J Aust* 2006;184(10 Suppl):S51–5.
3. Evans SM, Lowinger JS, Sprivilis PC, *et al*. Prioritizing quality indicator development across the healthcare system: identifying what to measure. *Intern Med J* 2009;39:648–54.
4. Genovitch-Richards J. Quality management organizational structures: history and trends. *J Healthc Qual* 1994;16:22–5.
5. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988;260:1743–8.
6. Mainz J. Defining and classifying clinical indicators for quality improvement. *Int J Qual Health Care* 2003;15:523–30.
7. NELA Project Team. Second Patient Report of the National Emergency Laparotomy Audit. 2016 www.nela.org.uk/reports.
8. www.PQIP.org.uk
9. <https://www.facs.org/quality-programs/acs-nsqip>
10. Copnell B, Hagger V, Wilson SG, *et al*. Measuring the quality of hospital care: an inventory of indicators. *Intern Med J* 2009;39:352–60.
11. Wollersheim H, Hermens R, Hulscher M, *et al*. Clinical indicators: development and applications. *Neth J Med* 2007;65:15–22.
12. Arling G, Kane RL, Lewis T, *et al*. Future development of nursing home quality indicators. *Gerontologist* 2005;45:147–56.
13. 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.
14. Haller G, Stoelwinder J, Myles PS, *et al*. Quality and safety indicators in anesthesia: a systematic review. *Anesthesiology* 2009;110:1158–75.
15. Australian Council on Healthcare Standards. Clinical indicators definitions. Available at www.achs.org.au
16. Anesthesia Quality Institute-Outcome definitions. Available at www.aqihq.org/.AQI_Outcomes_Definitions. 2015
17. Haller G, Laroche T, Clergue F. Morbidity in anaesthesia: today and tomorrow. *Best Pract Res Clin Anaesthesiol* 2011;25:123–32.
18. Grocott MP. Standardisation of perioperative outcome measures. *Eur J Anaesthesiol* 2015;32:83–4.
19. Myles PS, Grocott MP, Boney O, *et al*. Standardizing end points in perioperative trials: towards a core and extended outcome set. *Br J Anaesth* 2016;116:586–9.

20. Campbell SM, Braspenning J, Hutchinson A, *et al.* Research methods used in developing and applying quality indicators in primary care. *Qual Saf Health Care* 2002;11:358-64.
21. Mainz J. Developing evidence-based clinical indicators: a state of the art methods primer. *Int J Qual Health Care* 2003;15 Suppl 1:5i-11.
22. National library of Medicine. Abridged Index Medicus (AIM or "Core Clinical") Journal Titles. Available at <https://www.nlm.nih.gov/bsd/aim.html>
23. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376-80.
24. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol* 2003;56:1150-6.
25. Last J. *A dictionary of Epidemiology. A Handbook sponsored by the International Epidemiology Association*. Fourth edition. New York: Oxford University Press, 2001.
26. Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. Available at <http://training.cochrane.org/handbook>
27. Koroshetz W. A core set of trial outcomes for every medical discipline? *Br Med J* 2015; 350; Ioannidis JP. How to make more published research true. *PLoS Med* 2014;11:e1001747.
28. Boney O, Moonesinghe SR, Myles PS, *et al.* Standardizing endpoints in perioperative research. *Can J Anaesth* 2016;63:159-68.
29. Ioannidis JP. How to make more published research true. *PLoS Med* 2014;11:e1001747-85.
30. Lohr K. Committee to design a strategy for quality review and assurance. *Medicare: a strategy for quality assurance*. Washington, DC: National Academy Press, 1990:1.