Systematic reviews and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: mortality, morbidity and organ failure

Alexander I.R. Jackson*[1, 2], Oliver Boney[3,4], Rupert M. Pearse[5], Andrea Kurz[6], D. James Cooper[7,8], Wilton A. van Klei[9], Luca Cabrini[10], Timothy E. Miller[11], S. Ramani Moonesinghe [12,13,14], Paul S. Myles [15,16], Michael P.W. Grocott[1, 2], for The StEP-COMPAC Group

1. Perioperative and Critical Care Theme, NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK

2. Integrative Physiology and Critical Illness Group, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

3. Surgical Outcomes Research Centre, University College Hospital, London, UK

4. Health Services Research Centre, National Institute of Academic Anaesthesia, Royal College of Anaesthetists, London, UK

5. Faculty of Medicine & Dentistry, Queen Mary University of London, UK

6. Departments of General Anesthesiology and Outcomes Research, Cleveland Clinic, Cleveland, OH, United States

7. Australia New Zealand Intensive Care Research Centre, School of Public Health and Preventative Medicine, Monash University, Melbourne, Victoria, Australia

8. Department of Intensive Care and Hyperbaric Medicine, The Alfred, Melbourne, Victoria, Australia

9. Department of Anesthesia and Pain Management, Toronto General Hospital, University Health Network Toronto; Department of Anesthesiology and Pain Medicine, Temerty Faculty of Medicine, University of Toronto / Department of Anesthesiology and Intensive Care Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.

10. Department of Biology and Life Sciences, ASST Sette Laghi, Insubria University, Varese, Italy

11. Department of Anesthesiology, Duke University School of Medicine, Durham, NC.

12. Centre for Peri-Operative Medicine, Research Department for Targeted Intervention, University College London, UK

13. University College London / University College London Hospitals National Institute Health Research Biomedical Research Centre, London, UK

14. Department for Anaesthesia and Perioperative Medicine, University College London Hospitals, London, UK

15. Department of Anaesthesiology and Perioperative Medicine, Alfred Hospital, Melbourne, Australia

16. Department of Anaesthesiology and Perioperative Medicine, Monash University, Melbourne, Australia

*Corresponding Author: Alexander Jackson – alexander.jackson@soton.ac.uk, Anaesthesia and Critical Care Research Unit, CE93, Mailpoint24, ELevel, Centre Block, Southampton General Hospital, Tremona Road, Southampton SO16 6YD, UK

Abstract

BACKGROUND: Mortality, morbidity and organ failure are important and common serious harms following surgery. However, there are many candidate measures to describe these outcome domains. Definitions of these measures are highly variable, and validity is often unclear. As part of the International Standardised Endpoints in Perioperative Medicine (StEP) initiative, this study aimed to derive a set of standardised and valid measures of mortality, morbidity and organ failure for use in perioperative clinical trials.

METHODS: Three domains of endpoints (mortality, morbidity and organ failure) were explored through systematic literature review and a three-stage Delphi consensus process using methods consistently applied across the StEP initiative. Reliability, feasibility and patient-centeredness were assessed in round three of the consensus process.

RESULTS: A high level of consensus was achieved for two mortality time-points - 30 day and 1-year mortality – and these two measures are recommended. No organ failure endpoints achieved threshold criteria for consensus recommendation. The Clavien-Dindo Classification of complications achieved threshold criteria for consensus in round two of the Delphi process but did not achieve the threshold criteria in round three where it scored equivalently to the Post Operative Morbidity Survey. Clavien-Dindo therefore received conditional endorsement as the most widely used measure. No composite measures of organ failure achieved an acceptable level of consensus.

DISCUSSION: Both 30-day and 1-year mortality measures were recommended. No measure was recommended for organ failure. One measure (Clavien-Dindo) was conditionally endorsed for postoperative morbidity, but our findings suggest that no single endpoint offers a reliable and valid measure to describe perioperative morbidity that is not dependent on the quality of delivered care. Further refinement of current measures, or development of novel measures, of postoperative morbidity may improve consensus in this area.

KEYWORDS: perioperative; outcomes; mortality; morbidity; consensus; methodology; anaesthesia; surgery

Introduction

Mortality and survival are amongst the most widely used and readily understood clinical endpoints across all clinical research. Yet even these seemingly straightforward measures are inconsistently reported with multiple timepoints and definitions (e.g. all-cause/overall, cancer attributable, in-hospital) found in the literature.[1] Death following elective surgery is, thankfully, a relatively uncommon occurrence.[2] This necessitates alternative markers of serious (non-fatal) postoperative harm, commonly referred to as morbidity or complications. Such events are substantially more common than mortality[2] but are also conceptually more complicated to measure with varying severity, duration, and underlying pathology all being important. This variety amplifies the need for standardised outcome measures. The Standardised Endpoints in Perioperative Medicine (StEP) initiative aims to provide guidelines for outcome selection through an expert, consensus-based approach. This will help standardise reporting, aid in the comparison of studies and facilitate effective synthesis of evidence through systematic review and meta-analysis.[3] We report the work of the StEP-COMPAC theme group for mortality, morbidity and organ failure.

Methods

Overview of methods

Three separate systematic reviews were undertaken to identify outcome measures used for each of mortality, morbidity and organ failure. The results of these reviews were used to derive a long list of candidate endpoints, which were subsequently assessed using a 3-stage Delphi consensus process.

Literature Searches

Mortality

This review was conducted by a single author (OB). All randomised controlled trials, involving adult major surgery, published over a ten year time period between January 2005 and December 2014 in the four highest impact-factor 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) were reviewed. If mortality or survival was reported, the timepoint or other defining characteristics (e.g., in-hospital) was recorded and used as the basis for the long list of candidate endpoints.

Morbidty and Organ Failure

These two reviews were conducted by a single author (AJ). Because morbidity and organ failure are less widely reported than mortality a different, broader, search strategy was adopted to ensure a diverse range of possible

outcomes were captured. The search was conducted via MEDLINE using the Abridged Index Medicus, with an additional 12 journals of special relevance to perioperative medicine included (see Appendix A for full search strategy). Randomised controlled trials and observational studies, in adult subjects undergoing any surgical intervention, from 2007-2017 were considered. A common search strategy was used to select the above criteria, with an overall aim to maximise sensitivity, with differences in spelling and classification allowed for. Separate search terms for organ failure and morbidity/complications were then used (see Appendix A) to generate two separate reviews. For the morbidity search, any endpoint which reported morbidity or complications as a composite endpoint was included, while organ specific measures (e.g. cardiac morbidity) were excluded. For the organ failure search all measures reporting multiple organ failure (MOF) as a composite were considered. In cases where studies did not explicitly state this, they were included if the tools were originally designed as measures of MOF. Measures identified in these two searches were then considered for long-listing. Endpoints which were speciality, or procedure specific (e.g. Atlanta Classification for MOF in pancreatitis) were not taken forward to the Delphi process.

Delphi Consensus Process

The longlist of potential outcomes generated from the literature review was tabulated into a standardised StEP Delphi proforma. This proforma allowed working group members to review and rate the outcomes. For round 1 in addition to outcome definitions, the initial studies defining the outcome measure were listed, as was a summary of literature review findings. In subsequent rounds a summary of the previous rounds findings were listed. The process was co-ordinated by a single author (AJ) working with the UK's NIAA Health Services Research Centre.[4]

In round 1 (R1) of the Delphi process the StEP Mortality, Morbidity and Organ Failure working-group members (n=10) were asked to score each item on the preliminary list using an ordinal scale (1 to 9), with 1 to 3 labelled 'not that important or invalid', 4 to 6 labelled 'important but requires revision', and 7 to 9 labelled 'critical for inclusion' and were given the option to select 'unsure'. Participants were invited to add any other endpoints, definitions, or modifications to existing definitions that they believed should be considered further. Space was provided for free text comment in relation to each endpoint. At the end of R1, measures receiving a score of ≥ 7 from ≥50% of respondents, or a score of 9 (critical for inclusion) from any participant, were taken forward to round 2. If no measures within the sub-domain (mortality, morbidity and MOF) met the threshold then the highest scoring measure for that sub-domain was taken forward.

In round 2 (R2) of the Delphi process all members of the StEP working group (n=89) were asked to rate the endpoints brought forward from R1 using the same ordinal scale as R1. A summary of the R1 findings were added to the proforma including: reason for inclusion in R2, median score, percentage of responses ≥7 and either a summary or verbatim comments. Following the completion of R2 the results were tabulated and endpoints which received a score of ≥7 from ≥70% of participants, were taken forward to the final round.

The final round (R3) of the process again was limited to sub-group members (n=10). A modified version of the previous proforma was used, again with a summary of the previous round's (R2) results included. The modification allowed not only for the overall ordinal scale (1 to 9) but also ordinal scales (1-9) for each of face validity, content validity, reliability, feasibility and patient centeredness. These elements have been previously described[5] but may be summarised as:

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-centredness - Does the endpoint have 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 Reviews

The mortality literature review examined 412 trials reporting mortality/survival, from this the six timepoints in regular use were selected for long-listing. The process for the other two literature searches is outlined in Figures 1 and 2. These identified four composite measures of morbidity (Clavian-Dindo,[6] CCI,[7] Accordion,[8] and POMS[9]) and three of organ failure (MODS,[10] Denver[11] and SOFA[12]) for inclusion in the long list.

Delphi Consensus Development

The response rates for rounds 1, 2 and 3 were 60%, 74%, and 90% respectively. The overall results for each round are summarised in Table 1-4.

In R1 two endpoints from each of morbidity and mortality were discarded leaving two and four endpoints for each domain respectively. For organ failure no measure met the threshold to be carried forward, therefore the highest scoring endpoint, the Sequential Organ Failure Assessment (SOFA) score, was selected for inclusion. In addition, one group member suggested the quick SOFA (qSOFA) score for inclusion.

In R2, again, no organ failure measures met criteria for inclusion in the final round. In the morbidity domain a single endpoint (Clavien-Dindo) reached the threshold with a score of ≥ 7 from 73% of respondents. Much higher overall agreement rates were seen in the mortality domain with both 30-day and 1 year mortality scoring ≥ 7 from 92% and 85% of respondents respectively.

For R3 it was decided to carry forward two endpoints not meeting criteria in R2 (POMS and SOFA). This was done to provide additional narrative information and comment on these endpoints rather than as potential endpoints for recommendation. SOFA was carried forward as it represented the highest scoring organ failure measure, while POMS received a number of comments in support as well as suggested modifications to enhance its utility. As such further comment from an expert panel was felt to be beneficial.

The results of R3 demonstrate high levels of ongoing consensus for the mortality measures with strong scores for both overall acceptance (Table 1) and across the individual scoring domains (Table 2). No other measures reached an overall consensus threshold of ≥ 7 from ≥ 70% of respondents (Table 1).

Discussion

Three separate reviews of high impact literature were performed yielding a series of endpoints for consideration across the three domains of mortality, morbidity and organ failure. A three stage Delphi consensus process was subsequently completed to identify endpoints which should be recommended for the measurement of these three constructs in future perioperative medicine studies.

Mortality represents one of the most widely understood and reported outcomes in healthcare, but it is not without its challenges.[13] The review conducted for this study demonstrated heterogeneity in reporting, while variation also exists in collection and verification.[13] However, our Delphi process demonstrated a high level of consensus amongst experienced perioperative researchers for mortality measures with two time points meriting firm recommendation by this group. Both 30-day and 1-year mortality consistently scored highly in each round and across all domains in which they were considered. We therefore encourage researchers to consider these time-points when designing future studies. The combination of 30 days and 1 year offer both short and longer-term outcome data, respectively. However, throughout the process comments highlighting specific cases where other measures may be useful should be considered. For example, in-hospital mortality may be more feasible in locations where reliable out of hospital mortality data are unavailable.[14] We acknowledge that such cases exist but where possible we feel consistent reporting of 30-day and 1-year mortality will ensure a degree of consistency not currently seen in the literature. The 1-year mortality recommendation also harmonises with the "patient-centred outcomes" StEP group's support for reporting 1-year disability free survival[15] using the World Health Organisation's Disability Assessment Score 2.0 (WHODAS 2.0)[16]

The organ failure domain, did not produce such a level of consensus. In each of the three rounds, this was consistently the domain where contributors recorded the lowest scores, which consequently did not reach threshold criteria for consensus. The SOFA Score,[12] a widely used endpoint in the intensive care literature,[17] was reviewed in each round. Comments received during this process suggest that in the perioperative setting the relative rarity of multiple organ failure limits the utility of this endpoint. Furthermore, in R3 doubts were

expressed over the patient centredness and face validity of SOFA in the typical perioperative population. The authors suggest that this measure may have a valid place in certain high-risk cohorts or sub-groups where organ failure may be anticipated, but do not advocate its routine use in perioperative medicine studies. Indeed, based upon our Delphi process there does not appear to be a high level of support for the routine reporting of any composite measure of organ dysfunction. Other working groups within the StEP initiative have identified measures of individual organ dysfunction[18-21] and it may be that using such measures alone, or in combination, may be the best strategy in studies where particular organs are the focus of interest.

The morbidity domain represented perhaps the most challenging domain to explore and interpret. Mortality after elective surgery is, thankfully, a relatively rare event,[2,22,23] although more common after some major emergency surgery such as laparotomy or hip fracture.[24,25] Morbidity, or complications, after surgery is much more common, occurring between 10 and 30 times more frequently.[2,23] This is important to note, not only because of the additional statistical power for any given sample size (thereby enabling smaller and less costly trials), but also from a patient centered perspective: morbidity will affect a far greater number of patients in the perioperative period than mortality will. It also is associated with worse longer-term health outcomes, including survival[26] and health-related quality of life.[27] However, morbidity remains a challenging construct to measure, and this was reflected in the results of our Delphi process. The endpoint which consistently scored the highest was the Clavien-Dindo Classification.[6] This narrowly achieved our definition of consensus (score ≥ 7 from ≥ 70% of respondents) in R2 (73%) but fell short in R3 (55.6%). During the process, comments suggested a limitation is that the severity of complications in Clavien-Dindo is defined by the level of intervention required to manage them. This generates two principle limitations: firstly, it will to a greater or lesser extent be dictated by resource availability, of particular relevance in low and middle-income countries, and secondly may not reflect the true physiological severity of the morbidity experienced, particularly as a single score is used to reflect an entire inpatient stay. This perhaps explains why the lowest scoring sub-domain in round 3 for Clavien-Dindo was that of patient centeredness. However, Clavien-Dindo does have notable advantages. It was considered in R3 to be both reliable and feasible, which is reflected in its widespread use throughout the literature. Despite its limitations, it remains the most widely accepted measure of postoperative morbidity in our process. Given the importance of measuring perioperative morbidity we believe a conditional recommendation for its use is advocated. We would suggest that it is reported routinely in perioperative medicine studies but would encourage the use of alternative measures of morbidity, which seek to measure aspects not captured by Clavien-Dindo, where this is of particular interest.

The Post Operative Morbidity Survey (POMS)[9] failed to achieve the threshold criteria for consensus in R2 but did attract a substantial amount of commentary from our contributors. In addition, it scored more highly than Clavien-Dindo on patient-centredness, although it was less highly rated for both face and content validity. POMS takes a different approach to Clavien-Dindo in that it aims to capture any morbidity that might prevent hospital discharge, across 9 domains broadly mapping to organ systems. The principal limitation identified

from the comments in our process is that it may capture some transient morbidity of limited clinical significance. Efforts have been made to overcome this limitation, through the timing of the survey itself (e.g. near to the time of planned discharge at postoperative day 5 or 7) or by combining POMS with Clavien-Dindo to exclude minor morbidity.[28] It was for these reasons that we decided to carry the POMS forward to the final round of the Delphi process where it did not achieve threshold criteria for consensus but again attracted supportive commentary and overall equivalent number of scores ≥ 7 (55.6) to Clavien-Dindo. As a consequence, we do not endorse the routine reporting of this measure, but it was considered to have a role, and in the future, may be amenable to modification to increase its utility and acceptance.

The consensus results for Clavien-Dindo and POMS suggest that there may be a need to further refine the current measures of perioperative morbidity, or to develop novel measures. We propose that such an endpoint be based principally on patient physiology and pathology and focus on capturing morbidity identified as important by patients and clinicians alike. Defining morbidity by the healthcare interventions and processes delivered care, as is the case for Clavien-Dindo and POMS (at least in part), has the substantial limitation that the recorded incidence/prevalence of morbidity is in part dependent on the quality of delivered care.

This study has several limitations. Our three literature reviews all focused on high impact journals and, by each examining literature from a 10-year period, will favour longer established endpoints. We accept that some novel, less widely utilised, measures may have been initially missed. We also acknowledge the use of a single reviewer as a limitation. However, these are offset by the contribution of an extensive panel of international experts to suggesting additional endpoints and we believe important outcomes of interest will all have been captured. Our response rate in R1 was relatively low, however, we were able to re-engage the panel to reach credible consensus decisions, as evidenced by the increasing response rates through the process.

Conclusion

Three separate domains of endpoints have been considered: mortality, morbidity, organ failure. Each have been assessed by literature review and a 3-stage Delphi consensus process. A high level of consensus was achieved for two mortality endpoints and we recommend the routine reporting of 30-day and 1-year mortality, where this is achievable and relevant to the aims of the research. No endpoints for organ failure achieved threshold criteria for consensus; we therefore do not advocate the routine reporting of composite organ failure in perioperative medicine studies and would instead direct readers to the StEP recommendations for individual organ systems.[12-15] No measure of morbidity achieved threshold criteria for consensus in the final round (3). However, the Clavien-Dindo Classification achieved threshold criteria for consensus in round 2 and this was conditionally endorsed as the most widely accepted measure. However,

we believe at present no single endpoint offers a reliable valid measure to describe perioperative morbidity that is not dependent on the quality of delivered care.

Authors' Contributions

Study Concept: PSM

Protocol development: PSM, MPWG

Systematic reviews: OB, AIRJ

Management of Delphi rounds: AIRJ

Participation in Delphi survey: All Authors

First draft and revision of manuscript: MPWG, AIRJ

Critical review and revisions of the manuscript: All authors

Acknowledgements

Funding

Declaration of Interest

RP holds research grants and/or honoraria from Edwards Lifesciences, Intersurgical and GlaxoSmithkline.

MG is an elected council member of the Royal College of Anaesthetists, board chair of the National Institute of Academic Anaesthesia, and deputy-chair of the UK national Centre for Perioperative Care. MG has received unrestricted research funding from Edwards Lifesciences Ltd, Pharmacosmos Ltd and Sphere Medical Ltd. He has served on the medical advisory board of Sphere Medical Ltd and Edwards Lifesciences Ltd.

References

1. Grocott MPW, Dushianthan A, Hamilton MA, et al. Perioperative increase in global blood flow to explicit defined goals and outcomes following surgery. *Cochrane Database Syst Rev* 2012; **11**: CD004082International Surgical Outcomes Study group. Global patient outcomes after elective surgery: prospective cohort study in 27 low-, middle- and high-income countries. en. *Br. J. Anaesth.* **117,** 601–609 (Oct. 2016).

2. Myles PS, Grocott MPW, Boney O, Moonesinghe SR, COMPAC-StEP Group. Standardizing end points in perioperative trials: towards a core and extended outcome set. *Br J Anaesth* 2016; **116**: 586–9

3. NIAA Health Services Research Centre [Internet]. [cited 2022 May 20]. Available from: https://www.niaa-hsrc.org.uk/

4. 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**: 705–11

5. Dindo D, Demartines N, Clavien P-A. Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Ann Surg* 2004; **240**: 205–13

6. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien P-A. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg* 2013; **258**: 1–7

7. Strasberg SM, Linehan DC, Hawkins WG. The accordion severity grading system of surgical complications. *Ann Surg* 2009; **250**: 177–86

8. Bennett-Guerrero E, Welsby I, Dunn TJ, et al. The use of a postoperative morbidity survey to evaluate patients with prolonged hospitalization after routine, moderate-risk, elective surgery. *Anesth Analg* 1999; **89**: 514–9

9. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. *Crit Care Med* 1995; **23**: 1638–52

10. Moore FA, Sauaia A, Moore EE, Haenel JB, Burch JM, Lezotte DC. Postinjury multiple organ failure: a bimodal phenomenon. *J Trauma* 1996; **40**: 501–10; discussion 510–2

11. Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22**: 707–10

12. Eisenstein EL, Prather K, Greene SJ, et al. Death: The Simple Clinical Trial Endpoint. *Stud Health Technol Inform* 2019; **257**: 86–91

13. Meara JG, Leather AJM, Hagander L, et al. Global Surgery 2030: evidence and solutions for achieving health, welfare, and economic development. *Lancet* 2015; **386**: 569–624

14. Moonesinghe SR, Jackson AIR, Boney O, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine initiative: patient-centred outcomes. *Br J Anaesth* 2019; **123**: 664–70

15. World Health Organization. Measuring Health and Disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0. World Health Organization; 2010.

16. Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; **315**: 801–10

17. McIlroy DR, Bellomo R, Billings FT 4th, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: renal endpoints. *Br J Anaesth* 2018; **121**: 1013–24

18. Abbott TEF, Fowler AJ, Pelosi P, et al. A systematic review and consensus definitions for standardised end-points in perioperative medicine: pulmonary complications. *Br J Anaesth* 2018; **120**: 1066–79

19. Beattie WS, Lalu M, Bocock M, et al. Systematic review and consensus definitions for the Standardized Endpoints in Perioperative Medicine (StEP) initiative: cardiovascular outcomes. *Br J Anaesth* 2021; **126**: 56–66

20. Barnes J, Hunter J, Harris S, et al. Systematic review and consensus definitions for the Standardised Endpoints in Perioperative Medicine (StEP) initiative: infection and sepsis. *Br J Anaesth* 2019; **122**: 500–8

21. Noordzij PG, Poldermans D, Schouten O, Bax JJ, Schreiner FAG, Boersma E. Postoperative mortality in The Netherlands: a population-based analysis of surgery-specific risk in adults. *Anesthesiology* 2010; **112**: 1105–15

22. 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**: 249–57

23. Saunders DI, Murray D, Pichel AC, Varley S, Peden CJ, UK Emergency Laparotomy Network. Variations in mortality after emergency laparotomy: the first report of the UK Emergency Laparotomy Network. *Br J Anaesth* 2012; **109**: 368–75

24. Royal College of Physicians. National Hip Fracture Database annual report 2019. London: RCP; 2019.

25. Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. *Ann Surg* 2005; **242**: 326–41; discussion 341–3

26. Archer S, Pinto A, Vuik S, et al. Surgery, Complications, and Quality of Life: A Longitudinal Cohort Study Exploring the Role of Psychosocial Factors. *Ann Surg* 2019; **270**: 95–101

27. Wong DJN, Oliver CM, Moonesinghe SR. Predicting postoperative morbidity in adult elective surgical patients using the Surgical Outcome Risk Tool (SORT). *Br J Anaesth* 2017; **119**: 95–105
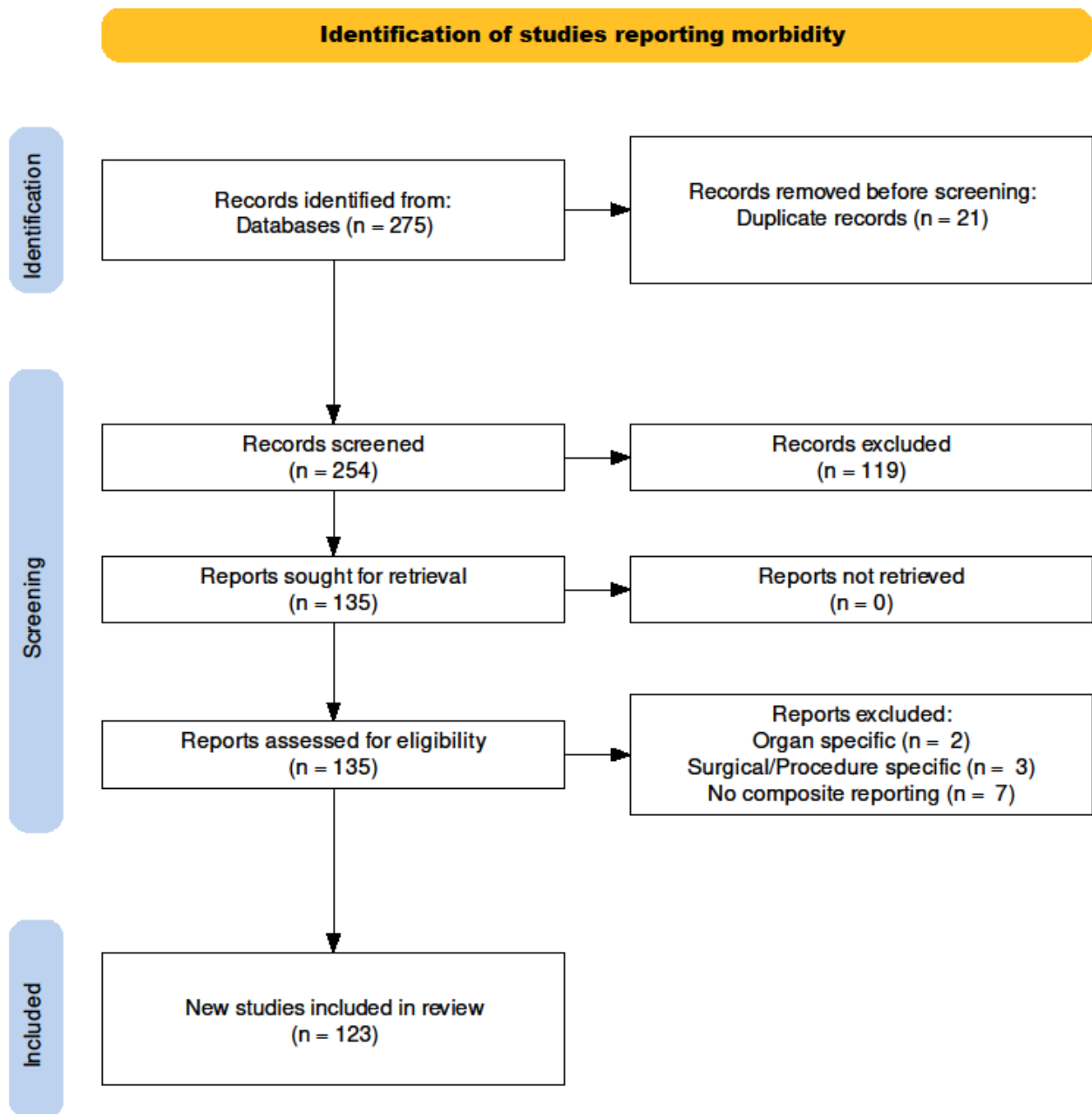
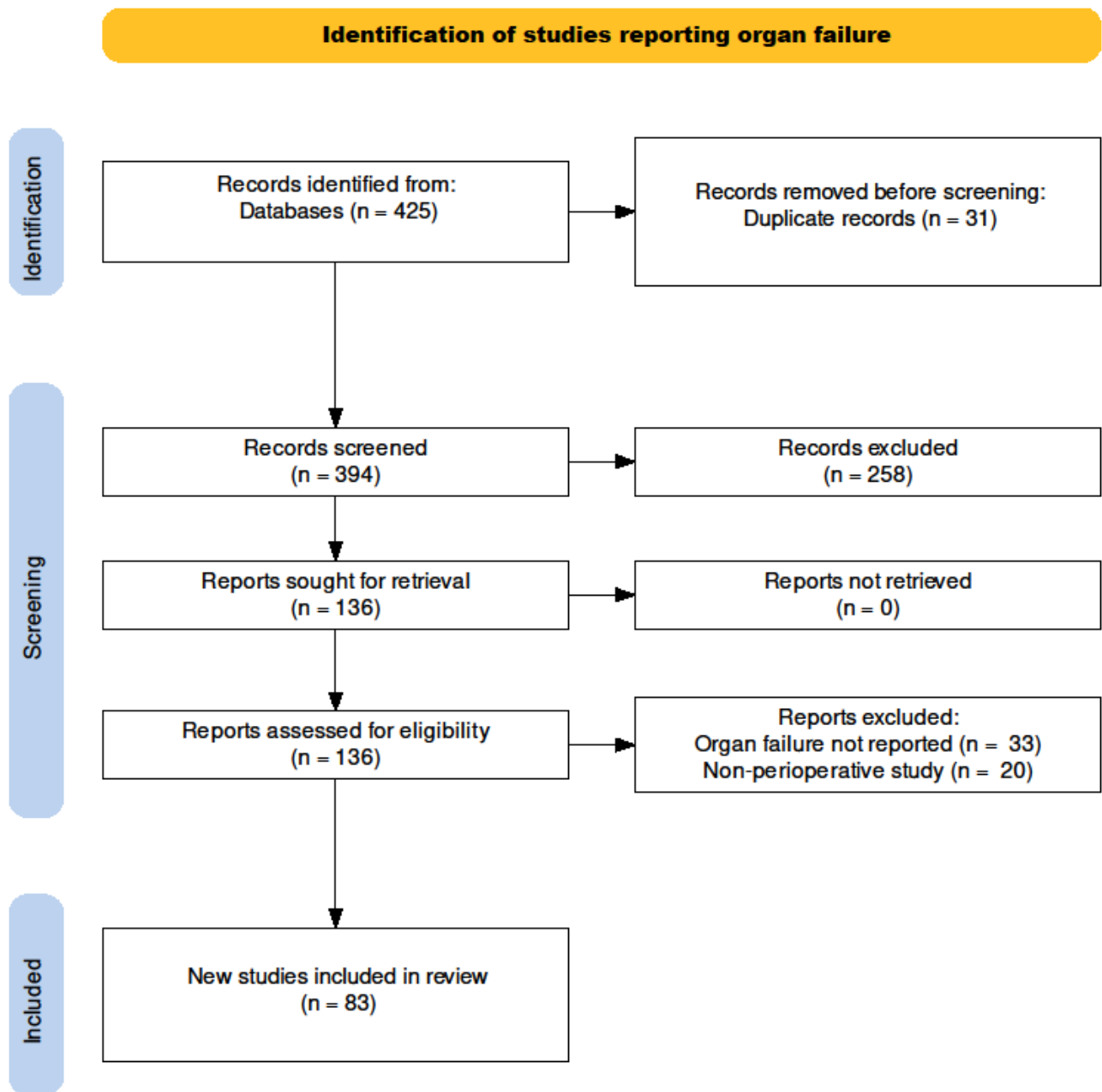**Figure 1: Flow Diagram of morbidity/complication search**

Figure 2: Flow Diagram of multi-organ failure search

Endpoint	Round 1			Round 2			Round 3		
	Unsure	Median score	%of scores $\geq$ 7	Unsure	Median score	%of scores $\geq$ 7	Unsure	Median score	%of scores $\geq$ 7
Mortality									
30 day	0	8.5	100	1	8	92	1	8	87.5
In hospital	0	7	66.7	1	7	65	-	-	-
1 year	0	7	100	4	8	85	2	8	100
90 day	0	7	50	2	7	67	-	-	-
28 day	0	4	0	-	-	-	-	-	-
60 day	0	4.5	16.7	-	-	-	-	-	-
Morbidity									
CD[1]	0	7.5	66.7	3	7	73	0	7	55.6
CCI[2]	0	4	0	-	-	-	-	-	-
Accordion	0	4.5	0	-	-	-	-	-	-
POMS[3]	0	6.5	50	3	6	38	0	7	55.6
Organ Failure									
Denver	0	4.5	0	-	-	-	-	-	-
MODS[4]	0	5.5	16.7	-	-	-	-	-	-
SOFA[5]	0	6	33.3	2	6	23	0	5	0
QSOFA[6]	-	-	-	4	6	27	-	-	-

[1]Clavien-Dindo Classification, [2]Comprehensive Complication Index, [3]Postoperative Morbidity Survey, [4]Multiple Organ Dysfunction Score, [5]Sequential Organ Failure Assessment, [6]Quick Sequential Organ Failure Assessment

Table 1: Results of the Delphi consensus process

Rating	30-Day Mortality					1-Year Mortality				
	Face Validity	Content Validity	Reliability	Feasibility	Patient centeredness	Face Validity	Content Validity	Reliability	Feasibility	Patient centeredness
High (%)	88.9	88.9	100	100	100	100	100	100	77.8	100
Moderate (%)	11.1	11.1	0	0	0	0	0	0	22.2	0
Low (%)	0	0	0	0	0	0	0	0	0	0

Table 2: Results of Round 3 Delphi Process (Mortality measures)

Rating	Clavien-Dindo					POMS[1]				
	Face Validity	Content Validity	Reliability	Feasibility	Patient centerednes	Face Validity	Content Validity	Reliability	Feasibility	Patient centerednes
High (%)	66.7	55.6	85.7	88.9	33.3	37.5	37.5	75	75	50
Moderate (%)	33.3	33.3	14.3	11.1	66.7	62.5	62.5	25	25	50
Low (%)	0	11.1	0	0	0	0	0	0	0	0

[1]Postoperative Morbidity Survey

Table 3: Results of Round 3 Delphi Process (Morbidity measures)

Rating	SOFA[1]				
	Face Validity	Content Validity	Reliability	Feasibility	Patient centeredness
High (%)	11.1	22.2	44.4	33.3.	0
Moderate (%)	66.7	55.6	44.4	44.4	66.7
Low (%)	22.2	22.2	11.1	22.2	33.3

[1]Sequential Organ Failure Assessment

Table 4: Results of Round 3 Delphi Process (Organ failure measures)