Standardizing endpoints in perioperative research

La standardisation des critères d'évaluation en recherche périopératoire

Oliver Boney, MBBS, MA · Suneetha R. Moonesinghe, MBBS, MD(Res) ·
Paul S. Myles, MPH, MD · Michael P. W. Grocott, MBBS, MD

Received: 14 September 2015 / Accepted: 11 November 2015 / Published online: 7 January 2016
© Canadian Anesthesiologists' Society 2016

Abstract *Measuring patient-relevant, clinically important, and valid outcomes is fundamental to the delivery of high-quality clinical care and to the innovation and development of such care through research. As surgical innovations become more complex and the burden of age and comorbidities in the surgical patient population continues to increase, understanding the benefits and harms of surgical interventions becomes ever more important. Nevertheless, we can understand only what we can adequately describe. Truly collaborative decision-making, delivery of safe effective care, and on-*

going quality improvement are also critically dependent on reliable valid measurement of patient-relevant and clinically important data. Attempts to describe the full spectrum of outcomes following surgery necessarily entail moving beyond the traditional endpoints of mortality and resource use towards more complex measures of morbidity, patient-reported outcomes, and functional status. Without standardization and consensus to guide the use of increasingly complex and nuanced endpoints, there is a real risk that perioperative research will become embroiled in a mire of inconsistent heterogeneous outcome measures that cannot be meaningfully compared and contrasted between trials or combined within meta-analyses. This would result in limiting the value of the research effort and depriving patients and clinicians of definitive answers. Collaboration in perioperative medicine—whether between institutions or across continents—has enormous potential to improve the value of research output. Standardizing endpoints for outcome measurement is fundamental to maximizing the quality of such collaboration and ensuring the impact of future perioperative research.

O. Boney, MBBS, MA · S. R. Moonesinghe, MBBS,
MD(Res) · M. P. W. Grocott, MBBS, MD
Surgical Outcomes Research Centre, University College
Hospital, London, UK

O. Boney, MBBS, MA · S. R. Moonesinghe, MBBS,
MD(Res) · M. P. W. Grocott, MBBS, MD
Health Services Research Centre, National Institute of Academic
Anaesthesia, Royal College of Anaesthetists, London, UK

P. S. Myles, MPH, MD
Department of Anaesthesia and Perioperative Medicine, Alfred
Hospital and Monash University, Melbourne, Australia

M. P. W. Grocott, MBBS, MD
Integrative Physiology and Critical Illness Group, Clinical and
Experimental Sciences, Faculty of Medicine, University of
Southampton, University Road, Southampton, UK

M. P. W. Grocott, MBBS, MD
Critical Care Research Area, Southampton NIHR Respiratory
Biomedical Research Unit, Southampton, UK

M. P. W. Grocott, MBBS, MD (✉)
Anaesthesia and Critical Care Research Unit, University
Hospital Southampton NHS Foundation Trust, Southampton, UK
e-mail: mike.grocott@soton.ac.uk

Résumé *Afin de fournir des soins cliniques de qualité supérieure et de promouvoir l'innovation et la mise au point de tels soins grâce à la recherche, il est essentiel de mesurer des critères d'évaluation qui soient à la fois pertinents pour le patient, importants d'un point de vue clinique et valables. Alors que les innovations chirurgicales deviennent de plus en plus complexes et que le fardeau du vieillissement et des comorbidités continue de s'accroître parmi les populations de patients chirurgicaux, il devient encore plus important de bien comprendre les risques et les avantages des diverses interventions chirurgicales. Toutefois, nous ne pouvons*

comprendre que ce que nous sommes capables de bien décrire. Une prise de décision véritablement collaborative, la prestation de soins aussi efficaces que sécuritaires et l'amélioration continue de la qualité dépendent toutes de façon cruciale d'une évaluation valable et fiable de données pertinentes au patient et importantes d'un point de vue clinique. Toute tentative de description de l'éventail complet de critères d'évaluation après une chirurgie implique forcément d'aller au-delà des critères d'évaluation conventionnels que sont la mortalité et l'utilisation des ressources, en se tournant vers des mesures plus complexes de la morbidité, des conséquences déclarées par le patient et de l'état fonctionnel. Si l'on ne parvient pas à standardiser ces mesures et à atteindre un consensus pour guider l'utilisation de critères d'évaluation de plus en plus complexes et nuancés, nous courons le risque réel que la recherche périopératoire s'embourbe dans un magma de mesures d'évaluation hétérogènes et incohérentes qui ne pourront être comparées de façon utile ou contrastées d'une étude à l'autre, ni combinées dans une méta-analyse. Ceci aurait pour conséquence de limiter la valeur des efforts de recherche et de priver les patients et les cliniciens de réponses claires. La collaboration en médecine périopératoire—que ce soit entre institutions ou par-delà les continents—recèle un énorme potentiel pour améliorer la valeur des résultats de recherche. La standardisation des critères d'évaluation pour mesurer les résultats de chirurgie est essentielle si l'on veut optimiser la qualité de telles collaborations et garantir l'impact des recherches périopératoires futures.

Introduction

"If you cannot measure it, you cannot improve it."

William Thomas, Lord Kelvin (1824-1907)

Why measuring perioperative outcomes is important

Reliable measurement and recording of outcomes after surgery should be integral to the delivery and development of high-quality surgical care. For consent to be truly informed and decision-making collaborative, surgical patients have a right to know the expected results of the procedure to which they are consenting. Providers of surgical services need to be able to evaluate their processes of care and resultant outcomes in order to benchmark their practice against other providers. Surgeons and the team they work with (including anesthesiologists) have a professional duty to show that their practice is safe and competent. Additionally, the financial and logistical

planning of surgical service delivery relies on robust process and outcome data, e.g., how long patients are likely to stay in hospital, which patients may require postoperative critical care, and the frequency of expected postoperative complications.

Why outcome measurement is important in perioperative research

Outcome measurement in perioperative research is equally fundamental. The only means by which clinical trials can reliably discriminate between beneficial, ineffective, and harmful interventions is by employing outcomes that are relevant to patients, clinically important, and valid. Moreover, the perioperative evidence base requires not only relevant and valid outcome measurements within individual trials but also consistency of outcome measurements between different trials. Such consistency facilitates comparing and contrasting results between studies as well as combining the findings into high-quality systematic reviews—i.e., the “gold standard” of evidence-based medicine. Heterogeneity of outcome measurements limits the quantitative pooling of data from multiple trials within meta-analyses. At best, this diminishes confidence in the pooled estimates of an intervention's effect.¹ At worst, it may preclude any quantitative pooling of data within a systematic review and thereby substantially undermine the value of individual trials and the utility of the combined evidence base.

Inconsistency in outcome reporting

The problem of inconsistent outcome reporting and its consequences has long been recognized. A large systematic review of stroke outcome measures in 2000 reported that “there is little consistency in the measurement of outcome in acute stroke trials, and this may complicate interpretation of the results and reduce the likelihood of detecting worthwhile drug effects”.² In the perioperative setting, a review of surgical outcome reporting in 2002 noted that “inconsistent complication reporting is common in hospitals and in the surgical literature”.³ Furthermore, a recent Cochrane systematic review of perioperative hemodynamic management noted that none of 31 included studies used the same set of postoperative morbidities.⁴

Lack of consensus regarding what outcomes to measure and how they should be defined has stimulated recent interest in standardizing endpoints and outcome measures for perioperative research.^{5,6} The potential benefits of researchers adopting standardized endpoints should be self-evident. Measuring outcomes in a variety of different ways makes it difficult or impossible to compare results between

trials. On the other hand, if identical criteria are adopted for a given outcome and used consistently across trials, data from different trials can be easily compared, contrasted, and combined in meta-analyses to provide a more precise estimate of the direction and magnitude of the true effect. The use of individual patient data (IPD) meta-analysis may help overcome some of the variation in methodology or outcome reporting between trials and thus reduce the heterogeneity of the data; however, the IPD approach is considerably more time and resource intensive than using aggregate trial data for systematic reviews.^{7,8} Consistent reporting of outcomes also facilitates improved understanding of the nuances of different trial results based on the nature of the clinical context, the patient population studied, and the intervention administered. Any uncertainty whether differences between studies are related to definition of outcomes or true clinical effects is minimized with consistent outcome reporting.

Core outcome measures and standardized endpoint reporting

Various initiatives have been introduced, including development of core outcome sets and standardized definitions for specific outcome measures, in an attempt to address the problems associated with inconsistent outcome measurement and reporting. The Core Outcome Measures in Effectiveness Trials (COMET) program⁹ is notable in this respect. The COMET program developed from initiatives to develop a core outcome set in rheumatology and cancer medicine. An important development against this background has been initiatives to combine the core outcomes set approach with standardized definitions in order to provide a consistent and comprehensive toolkit for investigators designing the clinical trials of the future.

In clinical trials, standardization of endpoints requires two essential elements—first, a defined core outcome set reported consistently across all trials, and second, consistent definitions (criteria) for individual outcome measures. These two elements may be combined to provide a menu of endpoints, with criteria, for each outcome domain, along with a standardized core outcome set. In turn, each outcome domain can include a hierarchy of consistently defined measures appropriately selected according to the level of detail and precision relevant to a particular trial, but with all trials consistently reporting the core outcome set.

The aim of this narrative review is to:

- summarize the challenges inherent in measuring and defining perioperative outcomes
- describe current efforts to standardize endpoints in perioperative care

- discuss the potential implications of standardized endpoints for perioperative research and clinical practice

The landscape of perioperative outcome measurement

Types of outcome measures

Outcome or process?

The quality of perioperative care may be measured in various ways. Some “outcomes” are in fact more accurately described as “process” measures, invoking the Donabedian model whereby both process measures (i.e., what we do or the actions involved in delivering health care) and outcome measures (i.e., the results or effects of those processes) may closely reflect the quality of care.¹⁰ Examples of process measures include length of hospital stay and completion of the WHO checklist, whereas 30-day mortality and postoperative myocardial infarction are consequences or outcome measures.

Clinician-described or patient-reported

The measure of interest may be described by clinicians or reported by patients. Patient-reported outcomes (e.g., rating of postoperative pain on a numerical rating scale) are considered “subjective” because they reflect patients’ perceptions, whereas clinician-described outcomes (e.g., the presence or absence of myocardial injury) are considered “objective” measures based on clinical evidence. Nevertheless, there is clearly a degree of subjectivity in assessing certain outcomes (e.g., the presence or absence of postoperative atelectasis). Moreover, patient-reported outcomes are, by definition, important to patients, whereas clinician-described outcomes may not be as important to patients.

The concept of “patient-centred” outcomes—i.e., focusing explicitly on outcomes that matter to patients—has recently been proposed to quantify postoperative recovery.¹¹ For example, an asymptomatic postoperative rise in troponin may not affect a patient’s recovery. For that reason, it would not constitute a patient-centred outcome despite being a predictor of future cardiac events and therefore an important outcome to clinicians. A recent review asserted that “we are now entering a new era in medicine where patient-centred outcomes will determine what constitutes medical success or failure, not only doctors’ perceptions of success.”¹²

Adverse events or recovery

Outcome measures may also focus on adverse events or postoperative recovery. Both approaches have strengths

and weaknesses. Adverse events can generally be observed clinically and/or confirmed from diagnostic tests, whereas recovery can be assessed only from patients using patient-reported outcome measures (PROMs) or indirect measures of functional capacity. Patient-reported outcome measures may be assessed using either specific questionnaires quantifying recovery¹³⁻¹⁷ or more general questionnaires developed for evaluating health-related quality of life (HRQL).¹⁸ Functional capacity is generally assessed by a standardized measure of physical function (e.g., six-minute walk test).¹⁹

While PROMs, by definition, are patient-centred and may reveal important sequelae following surgery that measures of adverse events fail to capture, PROMs instruments require specific psychometric evaluation and validation. New PROM tools require validation against existing tools, and similarly, their reliability—i.e., the extent to which a test provides the same output given the same input—may need formal testing. The feasibility of a particular measure also requires evaluation. Without the necessary resources and expertise for their use, the utility of PROMs in practice is clearly limited. The use of outcome measurement tools that have not been appropriately tested is common in perioperative research²⁰ and provides another justification for developing standardized endpoints.

Notwithstanding that measures of adverse events might therefore appear to be a simpler and more objective approach for assessing outcomes than patient-reported or indirect markers of recovery, they do have limitations. After most types of surgery, in-hospital mortality or mortality within 30 days occurs too infrequently to be an adequate description of perioperative outcomes. Consequently, perioperative morbidity, a more common outcome, has become the focus of perioperative outcome measurement. Perioperative morbidity has additional implications for the use of healthcare resources and is increasingly recognized as a predictor of long-term outcome.^{21,22} Recent characterization of the high-risk surgical population has also steered attention towards perioperative morbidity.²³ Although this population comprises only 10-15% of all surgical patients, it accounts for over 80% of postoperative complications and resource costs. Nonetheless, morbidity endpoints may also fall short in describing outcomes after surgery, particularly beyond the immediate perioperative period. Recent evidence suggests that long-term recovery may take months to years—at any rate, significantly longer than the time frames examined in most perioperative research.²⁴⁻²⁶ Hence, there is a growing interest in short-term and longer term recovery endpoints to quantify the overall success of surgery.²⁷⁻²⁹

Other controversies persist concerning measurement of adverse events. First, the definitions used for specific complications are frequently inconsistent. A review of

surgical adverse events in 2001 found 41 different definitions and 13 grading scales for surgical-site infections among the 82 studies included in the review.³⁰ The recent debate regarding definitions of perioperative myocardial injury is likewise ongoing.^{31,32} Even seemingly simple binary constructs such as mortality may be presented in a variety of ways.^{33,34} Such issues clearly suggest that standardized endpoints would greatly facilitate comparison of data between trials.

Second, grading the severity of adverse events is problematic. Although distinguishing between trivial and life-threatening complications may be informative, the various systems developed for quantifying severity,³⁵⁻³⁸ e.g., the Clavien-Dindo classification,³⁵ all suffer from a degree of inherent subjectivity. Severity is variously based on the degree of physiological derangement, its duration, or the invasiveness of treatment required; however, the long-term sequelae—i.e., the impact on a patient's quality of life and life expectancy—are arguably more important criteria for judging severity of complications to patients.

Composite outcomes

Morbidity outcomes may be reported singly or combined to give composite outcomes, such as major adverse cardiac events. Composite measures have two principal advantages. First, they can increase the power of a study to detect differences between groups. Combining several individually rare outcomes into one composite outcome increases the event rate and thus reduces the sample size required to detect a significant difference between groups. Second, amalgamating all separate clinically important outcomes—e.g., composite outcome of death, dependency, and poor neurological function, as is commonly reported in stroke trials—provides a succinct quantitative overview of the overall benefit of an intervention.

However, composite outcomes obscure the detail of the individual components and may be misleading if the component outcomes are not broadly equivalent in both severity and incidence. If one complication within the composite outcome occurs much more frequently than the others, the overall rate of the composite outcome will be skewed towards the rate of that particular complication.³⁹ Meanwhile composite outcomes are patently impossible to interpret in the context of the existing literature if their component inputs vary between trials.⁴⁰

Morbidity scores

Postoperative morbidity scores have been developed to quantify and compare overall postoperative morbidity for different procedures. These include the Postoperative Morbidity Survey (POMS),^{41,42} Cardiac Postoperative

Morbidity Score (C-POMS),⁴³ and Comprehensive Complication Index.^{44,45} While a robust system for quantifying overall postoperative morbidity would appear advantageous, such systems also pose challenges. They may be cumbersome to administer and require specific training in their use, which limits their utility. They require thorough validation and share the weakness of other composite outcomes, i.e., they may mask important differences in rates of individual specific postoperative morbidities. The POMS and C-POMS, in particular, were validated as means of detecting morbidity that would prevent hospital discharge. As healthcare systems and processes change, such measures may require refinement.

Resource use measures

An alternative approach is the use of resource consumption measures (e.g., critical care and hospital length of stay and readmission rates after surgery). Such measures are undeniably important in economic analyses of perioperative care. These measures are not only readily accessible from hospital data but they are also patient-centred, since patients generally want to return home from hospital as soon as possible. Nevertheless, resource use is only a proxy for clinical outcome. Despite the abundant evidence linking postoperative complications with increased length of stay and higher costs,⁴⁶⁻⁵⁰ resource use is also affected by several hidden factors (e.g., availability of community-based support, clinician behaviour, and hospital policy) that may limit the reliability—in different institutions or in the same institution over time—of measures of resource use as markers of perioperative clinical outcomes.

To avoid these potential confounders, trials may report “time to medical fitness for discharge” as a measure of overall clinical outcome.^{51,52} Although this endpoint tells us little about the clinical concerns that may have delayed fitness for discharge, it is arguably a good marker of the overall early success of surgery and perioperative care. Nevertheless, its utility as an outcome measure presupposes that different hospitals (or clinical teams) have similar criteria for considering a patient medically fit for discharge. Therefore, agreement on a consistent threshold and standardized definition for “medically fit for discharge” is a clear prerequisite for meaningful interpretation.

Patient-reported experience measures (PREMs)

Finally, overall patient experience or satisfaction with care may be an outcome of interest. The importance of patient satisfaction after surgery and/or anesthesia has been

increasingly recognized both in its own right and as a metric for quality of care.⁵³⁻⁵⁵ Furthermore, there is some evidence linking positive patient experience and clinical outcome.⁵⁶⁻⁵⁸ Concerns have arisen, however, over the use of non-validated tools to measure patient experience despite the availability of well-validated instruments.^{59,60} Without an agreed standard to guide investigators, the use of untested instruments to assess perioperative patient satisfaction is likely to continue.

In summary, regardless of the approach to perioperative outcome measurement, the use of complex outcome measures, such as PROMs, or composite morbidity endpoints greatly increases the scope for variability between trials. This makes it all the more compelling to standardize endpoints in perioperative care. Furthermore, perioperative outcome measures must be valid, reliable, and pragmatic (i.e., feasible) based on their specific contexts and populations. Investigators need consensus-based guidance that will help them determine and precisely define the outcome measures as well as establish time points to record the outcome measures. As would be expected, making such decisions on an arbitrary basis leads to inconsistency and random variation in outcome reporting.

Core outcomes and standardization initiatives

Thus far, efforts to improve consistency of outcome measurement in perioperative research have taken two approaches. The first approach involves determining the most appropriate “outcome domains” to describe the perioperative care (i.e., what to measure). Not all outcome domains will be relevant for all perioperative trials; however, establishing a “core” set of perioperative outcomes would facilitate formulating reliable comparisons and combinations of data from trials that report those outcomes, thus enabling their inclusion in systematic reviews and allowing investigators to increase the value of their studies.

The second approach involves agreeing on definitions for specific endpoints (i.e., standardized criteria regarding how to measure) for each outcome domain, e.g., myocardial injury after surgery, to ensure all trials use standardized definitions for reporting. Standardization would improve consistency between trials, reduce the use of non-validated endpoints, and ensure the use of precise widely accepted definitions for specific endpoints that are often defined only vaguely, if at all, in the current literature. Standardized endpoints and core outcome measures may of course be combined, providing researchers with clear consensus-based guidance on which outcomes should be reported and how such outcomes should be defined.

Core outcome sets

The concept of core outcome sets initially grew from research on rheumatoid arthritis during the 1990s. The 1992 Outcome Measures in Rheumatology Clinical Trials (OMERACT) conference developed from increasing recognition that assessing the impact of interventions was impossible without a consensus on what outcomes should be measured and how they should be defined. An agreement on a core outcome set for rheumatology trials was reached at the conference.⁶¹ The OMERACT group has since developed and validated core outcome sets for several rheumatological conditions and has pioneered methodology for choosing measurement instruments via its OMERACT “filter”.⁶²

Core outcome sets are increasingly being developed in a wide range of medical disciplines, from eczema to colorectal cancer.^{63,64} The Core Outcome Measures in Effectiveness Trials (COMET) initiative was established in 2010 to meet this growing need by “bringing together people interested in the development and application of agreed standardized sets of outcomes”.⁹ A related aim of the COMET initiative is to guide the methodology of core set development, which has itself become a subject of considerable research interest.⁶⁵ While no specific guidelines have been published to date, a recent systematic review of methods used in developing core outcome sets identified three important principles, namely, involving all stakeholders (i.e., patients and carers as well as clinicians and researchers) at all stages, achieving widespread consensus (usually via some form of Delphi process), and methodological transparency.⁶⁶

Core outcome measures for perioperative and anesthetic care (COMPAC)

An initiative to develop a core outcome set for trials in perioperative medicine and anesthesia is currently underway.⁶ Its aim is to develop a basic “core set” of outcomes for reporting in all perioperative trials without placing any restriction on the reporting of other more specific outcomes in particular studies. The methodology for COMPAC is based on the COMET initiative’s recommendations. A group of stakeholders (comprising perioperative clinicians, patients, carers, and researchers) is first convened, and then key steps are taken to complete a “long list” of all relevant outcome measures. First, a comprehensive literature search is performed to describe existing perioperative outcome measurements, followed by stakeholder consultation exercises to identify any other potentially relevant outcomes. A Delphi process is then utilized through which stakeholders select a shortlist of candidate outcome measures, and finally, a consensus

process is employed to reach agreement on the core outcome set.

European Society of Anaesthesiology (ESA)/European Society of Intensive Care Medicine (ESICM) joint task force standards

The challenge of standardizing endpoints in perioperative research has recently been addressed by a joint task force of the ESA and the ESICM.⁵ This represents the first international collaboration with plans to develop standardized endpoint definitions, with the stated aim of “providing a methodological standard for use in large pragmatic clinical studies designed to improve patient outcomes after surgery”.

The task force, comprising 12 perioperative experts with diverse backgrounds, conducted a literature review of perioperative outcome assessment. They then discussed the evidence base to reach agreement on a standardized definition for 22 pre-specified perioperative complications and four composite outcomes, along with a simple severity grading and recommendations for measuring HRQL after surgery. The standard definitions produced were deliberately straightforward and user-friendly, reflecting their intended use in large pragmatic trials where ease of data collection is an important consideration. Similarly, the task force incorporated existing definitions already widely used in audit and research, no doubt mindful that new definitions might further add to the confusion in selecting perioperative outcome endpoints. Severity grading was consistently defined throughout and based on the degree of anticipated harm and need for clinical intervention.

The task force discouraged measures of resource use, considering them “unreliable surrogate markers of clinical outcome because they are affected by hospital and healthcare policy as well as by clinician behaviour”. Nevertheless, they endorsed the measurement of recovery after surgery, proposing the quality of recovery (QoR)-15 instrument as a suitable standard,¹⁷ and emphasized the importance of measuring HRQL following surgery. Then again, none of the suggested HRQL endpoints were developed for postoperative use except the Post-operative Quality of Recovery Scale (PQRS).⁶⁷ Since PQRS has been validated only up to three months postoperatively, the task force also highlighted the need for new instruments to assess long-term HRQL after surgery.

While the ESA/ESICM endeavour represented the first international effort to standardize outcome definitions for perioperative research, the recommendations were not presented as a final solution. The task force acknowledged that the process had strengths and limitations and that “further work may improve the definitions provided and broaden the scope”.

Nonetheless, this international collaboration not only highlights the current lack of consensus-based standards in perioperative outcome measurement but also provides a useful, patient-relevant, clinically important, and precisely defined candidate set of standardized endpoints for perioperative research.

BJA symposium “defining perioperative endpoints” (June 2015)

A symposium organized by Monash University and sponsored by the *British Journal of Anaesthesia* was held in June 2015 at the Collaborative Clinical Trials in Anaesthesia Conference at the Monash Centre in Prato Italy. The symposium, chaired by Paul Myles (Australia) and Mike Grocott (UK), brought together over 50 internationally recognized experts in perioperative research to discuss the challenges of developing standardized endpoints for perioperative clinical trials. The overall aim of achieving a broad internationally recognized consensus on endpoint measurement involves a multi-step collaborative development process incorporating themes from both the COMPAC initiative and the ESA/ESICM endpoints workstreams.

Discussions at the conference focused first on identifying major themes or outcome domains that require standardized endpoints and then on forming expert working groups for each domain identified. Each working group will involve four to eight experts from around the world whose task will be to review the existing literature and then propose candidate standardized endpoints for their respective outcome domains. Consensus on the standardized endpoints for each domain will then be sought across all participants, likely through a modified Delphi process. An overall working party will be established to coordinate the COMPAC and STandardized EndPoints in Perioperative care (STEPP) processes.

This endeavour to develop standardized endpoints will thus build on the recommendations from the ESA/ESICM initiative as well as inform the work of the COMPAC initiative towards a core perioperative outcome set. Uniquely among these initiatives, the latter stages of COMPAC/STEPP will include patients and carers in deciding which outcomes warrant inclusion in a core set for all perioperative trials. The results of COMPAC/STEPP will be presented at the 16th World Congress of Anaesthesiologists in Hong Kong in August 2016.

Implications of standardized endpoints

The development of core outcomes and standardized endpoints has considerable implications for perioperative

research. Methodological standards in research (to minimize the risk of bias) are universally recognized, and explicit standards exist for judging methodological quality.^{68–70} Similarly, standards have existed for many years for reporting different types of trials.^{71–73} These frameworks exist to promote high-quality research and to ensure its accurate dissemination; however, to date, there is a lack of objective standards or guidelines for outcome measurement in perioperative research.

As mentioned previously, several medical specialties are considerably ahead of perioperative medicine in this regard.^{61,65,74} The examples of rheumatoid arthritis and pain medicine neatly illustrate the difficulties in measuring complex or multidimensional outcomes without standardized definitions. Perioperative medicine research has clearly reached a level of complexity that also warrants standardized outcome measurement. Now that we no longer consider mortality or length of stay as adequate markers of surgical outcomes,⁷⁵ more complex multidimensional outcome measures are increasingly used to quantify overall postoperative morbidity and recovery. Reporting such endpoints mandates agreement on how they should be measured. Without explicit standards among the research community, definitions for outcome measurement are doomed to remain arbitrary.

The potential benefits of standardizing outcome measurement are particularly significant in perioperative medicine, owing to the paucity of large multinational randomized-controlled trials. As long as smaller scale trials remain the norm in perioperative research, systematic reviews incorporating such trials will continue to form the highest level of evidence to inform practice. Nevertheless, the strength of conclusions drawn in a systematic review depends on the similarity of the studies included.⁷⁶ The significant heterogeneity in methodology, sample populations, and outcome measurement prevents drawing such firm conclusions, which largely defeats the purpose of systematic reviews. Consensus on core outcome measures and standardized endpoints within the field of perioperative research would thus overcome a significant obstacle to conducting useful systematic reviews and meta-analyses.

Conclusions

Measuring outcomes that are patient-relevant, clinically important, and valid is fundamental to the delivery of high-quality clinical care and to innovation and development of such care through research. As surgical innovations become more complex and the burden of age and comorbidities in the surgical patient population continues to increase, understanding the benefits and harms of surgical interventions becomes ever more important.

Nevertheless, we can understand only what we can adequately describe. Truly collaborative decision-making, delivery of safe effective care, and ongoing quality improvement are all critically dependent on reliable valid measurement of patient-relevant and clinically important data. Attempts to describe the full spectrum of outcomes following surgery necessarily entail moving beyond the traditional endpoints of mortality and resource use towards more complex measures of morbidity, patient-reported outcomes, and functional status.

Without standardization and consensus to guide the use of increasingly complex and nuanced endpoints, there is a real risk that perioperative research will become embroiled in a mire of inconsistent heterogeneous outcome measures that cannot be meaningfully compared and contrasted between trials or combined within meta-analyses. This will result in limiting the value of the research effort and in depriving patients and clinicians of definitive answers. Collaboration in perioperative medicine—whether between institutions or across continents—has enormous potential to improve the value of research outputs. Standardizing endpoints for outcome measurement is fundamental to maximizing the quality of such collaboration and ensuring the impact of future perioperative research.

Author contributions: Oliver Boney contributed substantially to all aspects of this manuscript, including conception and design; acquisition, analysis, and interpretation of data and drafting the article. Ramani R. Moonesinghe, Paul S. Myles and Mike P.W. Grocott contributed substantially to the conception and design of the manuscript and drafting of the article.

Acknowledgements The authors gratefully acknowledge the input of colleagues attending the *British Journal of Anaesthesia* funded workshop on “Defining Peri-operative Endpoints” (Monash University Prato Centre, June 2015) into refining some of the ideas presented in this manuscript.

Competing interests Paul S. Myles and Michael P.W. Grocott are jointly chairing the Core Outcomes in Perioperative Care (COMPAC) and STandardized EndPoints in Perioperative Care (STEPP) working group. Paul S. Myles is an editor of the *British Journal of Anaesthesia* and an NHMRC Practitioner Fellow. Michael P.W. Grocott is director of the National Institute of Academic Anaesthesia (NIAA) Health Services Research Centre and is funded in part from the British Oxygen Company Chair of the Royal College of Anaesthetists, awarded by the NIAA. Michael P.W. Grocott also serves on the working group establishing the NIAA Clinical Trials Network and leads the Fit-4-Surgery research collaboration and the Xtreme Everest Oxygen Research Collaboration. Suneetha R. Moonesinghe is deputy director of the NIAA Health Services Research Centre and Director of the UCLH NIHR Biomedical Research Centre’s Surgical Outcomes Research Centre (SOuRCe). Suneetha R. Moonesinghe is funded as an Improvement Science Fellow by the Health Foundation. Some of this work was undertaken at University College Hospital London NHS Foundation Trust—UCL NIHR Biomedical Research Centre, which received a portion of funding from the UK Department of Health Research Biomedical Research Centres funding scheme. Some of this work was undertaken at University Southampton NHS

Foundation Trust - University of Southampton NIHR Respiratory Biomedical Research Unit, which received a portion of funding from the UK Department of Health Research Biomedical Research Units funding scheme. All funding was unrestricted. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. The Cochrane Collaboration, 2011. Available from URL: <http://handbook.cochrane.org> (accessed September 2015).
2. Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. *Stroke* 2000; 31: 1429-38.
3. Martin RC 2nd, Brennan MF, Jaques DP. Quality of complication reporting in the surgical literature. *Ann Surg* 2002; 235: 803-13.
4. Grocott MP, 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: CD004082.
5. Jammer I, Wickboldt N, Sander M, et al. Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions: a statement from the ESA-ESICM joint taskforce on perioperative outcome measures. *Eur J Anaesthesiol* 2015; 32: 88-105.
6. Health Services Research Centre of the National Institute of Academic Anaesthesia (Royal College of Anaesthetists); Surgical Outcomes Research Centre (SOuRCe) at University College Hospital, London; Grocott MP, Myles PS, Moonesinghe SR, Boney OC. Core Outcome Measures in Perioperative and Anaesthetic Care (COMPAC). Available from URL: www.comet-initiative.org/studies/details/632 (accessed September 2015).
7. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010; 340: c221.
8. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof* 2002; 25: 76-97.
9. COMET Initiative. Core Outcome Measures in Effectiveness Trials (COMET) Initiative. Available from URL: www.comet-initiative.org (accessed September 2015).
10. Donabedian A. The quality of care. How can it be assessed? *JAMA* 1988; 260: 1743-8.
11. Myles PS. Meaningful outcome measures in cardiac surgery. *J Extra Corpor Technol* 2014; 46: 23-7.
12. Kalkman CJ, Kappen TH. Patient-centered endpoints for perioperative outcomes research. *Anesthesiology* 2015; 122: 481-3.
13. Browne J, Jamieson L, Lewsey J, et al. Patient Reported Outcome Measures (PROMs) in Elective Surgery. Report to the Department of Health, London School of Hygiene and Tropical Medicine. Department of Health, U.K.; 2007. Available from URL: https://www.lshtm.ac.uk/php/departments/healthservices/researchandpolicy/assets/proms_report_12_dec_07.pdf (accessed September 2015).
14. Lee L, Mata J, Augustin BR, et al. A comparison of the validity of two indirect utility instruments as measures of postoperative recovery. *J Surg Res* 2014; 190: 79-86.
15. Herrera FJ, Wong J, Chung F. A systematic review of postoperative recovery outcomes measurements after ambulatory surgery. *Anesth Analg* 2007; 105: 63-9.

16. Gornall BF, Myles PS, Smith CL, et al. Measurement of quality of recovery using the QoR-40: a quantitative systematic review. *Br J Anaesth* 2013; 111: 161-9.
17. Stark PA, Myles PS, Burke JA. Development and psychometric evaluation of a postoperative quality of recovery score: the QoR-15. *Anesthesiology* 2013; 118: 1332-40.
18. Lee L, Elfassy N, Li C, et al. Valuing postoperative recovery: validation of the SF-6D health-state utility. *J Surg Res* 2013; 184: 108-14.
19. Moriello C, Mayo NE, Feldman L, Carli F. Validating the six-minute walk test as a measure of recovery after elective colon resection surgery. *Arch Phys Med Rehabil* 2008; 89: 1083-9.
20. Liu SS, Wu CL. The effect of analgesic technique on postoperative patient-reported outcomes including analgesia: a systematic review. *Anesth Analg* 2007; 105: 789-808.
21. 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.
22. Moonesinghe SR, Harris S, Mythen MG, et al. Survival after postoperative morbidity: a longitudinal observational cohort study. *Br J Anaesth* 2014; 113: 977-84.
23. Pearse RM, Harrison DA, James P, et al. Identification and characterisation of the high-risk surgical population in the United Kingdom. *Crit Care* 2006; 10: R81.
24. Lawrence VA, Hazuda HP, Cornell JE, et al. Functional independence after major abdominal surgery in the elderly. *J Am Coll Surg* 2004; 199: 762-72.
25. Feldman LS, Kaneva P, Demyttenaere S, Carli F, Fried GM, Mayo NE. Validation of a physical activity questionnaire (CHAMPS) as an indicator of postoperative recovery after laparoscopic cholecystectomy. *Surgery* 2009; 146: 31-9.
26. Tran TT, Kaneva P, Mayo NE, Fried GM, Feldman LS. Short-stay surgery: what really happens after discharge? *Surgery* 2014; 156: 20-7.
27. Carli F, Mayo N. Measuring the outcome of surgical procedures: what are the challenges? *Br J Anaesth* 2001; 87: 531-3.
28. Shulman MA, Myles PS, Chan MT, McIlroy DR, Wallace S, Ponsford J. Measurement of disability-free survival after surgery. *Anesthesiology* 2015; 122: 524-36.
29. Lee L, Tran T, Mayo NE, Carli F, Feldman LS. What does it really mean to "recover" from an operation? *Surgery* 2014; 155: 211-6.
30. Bruce J, Russell EM, Mollison J, Krukowski ZH. The measurement and monitoring of surgical adverse events. *Health Technol Assess* 2001; 5: 1-194.
31. Botto F, Alonso-Coello P, Chan MT, et al. Myocardial injury after noncardiac surgery: a large, international, prospective cohort study establishing diagnostic criteria, characteristics, predictors, and 30-day outcomes. *Anesthesiology* 2014; 120: 564-78.
32. Khan J, Alonso-Coello P, Devereaux PJ. Myocardial injury after noncardiac surgery. *Curr Opin Cardiol* 2014; 29: 307-11.
33. Hu Y, McMurtry TL, Wells KM, Isbell JM, Stukenborg GJ, Kozower BD. Postoperative mortality is an inadequate quality indicator for lung cancer resection. *Ann Thorac Surg* 2014; 97: 973-9.
34. Mayo SC, Shore AD, Nathan H, et al. Refining the definition of perioperative mortality following hepatectomy using death within 90 days as the standard criterion. *HPB (Oxford)* 2011; 13: 473-82.
35. Dindo D, Demartines N, Clavien PA. 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.
36. Sugawara Y, Tamura S, Makuuchi M. Systematic grading of surgical complications in live liver donors. *Liver Transpl* 2007; 13: 781-2.
37. Strasberg SM, Linehan DC, Clavien PA, Barkun JS. Proposal for definition and severity grading of pancreatic anastomosis failure and pancreatic occlusion failure. *Surgery* 2007; 141: 420-6.
38. Clavien PA, Sanabria JR, Strasberg SM. Proposed classification of complications of surgery with examples of utility in cholecystectomy. *Surgery* 1992; 111: 518-26.
39. Ross S. Composite outcomes in randomized clinical trials: arguments for and against. *Am J Obstet Gynecol* 2007; 196: 119.e1-6.
40. Myles PS, Devereaux PJ. Pros and cons of composite endpoints in anesthesia trials. *Anesthesiology* 2010; 113: 776-8.
41. Grocott MP, Browne JP, Van der Meulen J, et al. The Postoperative Morbidity Survey was validated and used to describe morbidity after major surgery. *J Clin Epidemiol* 2007; 60: 919-28.
42. Davies SJ, Francis J, Dilley J, Wilson RJ, Howell SJ, Allgar V. Measuring outcomes after major abdominal surgery during hospitalization: reliability and validity of the Postoperative Morbidity Survey. *Perioper Med (Lond)* 2013; 2: 1.
43. Sanders J, Keogh BE, Van der Meulen J, et al. The development of a postoperative morbidity score to assess total morbidity burden after cardiac surgery. *J Clin Epidemiol* 2012; 65: 423-33.
44. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien PA. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. *Ann Surg* 2013; 258: 1-7.
45. Slankamenac K, Nederlof N, Pessaux P, et al. The comprehensive complication index: a novel and more sensitive endpoint for assessing outcome and reducing sample size in randomized controlled trials. *Ann Surg* 2014; 260: 757-62; discussion 762-3.
46. Flynn DN, Speck RM, Mahmoud NN, David G, Fleisher LA. The impact of complications following open colectomy on hospital finances: a retrospective cohort study. *Perioper Med (Lond)* 2014; 3: 1.
47. Knechtle WS, Perez SD, Medbery RL, et al. The association between hospital finances and complications after complex abdominal surgery: deficiencies in the current health care reimbursement system and implications for the future. *Ann Surg* 2015; 262: 273-9.
48. Ebm C, Cecconi M, Sutton L, Rhodes A. A cost-effectiveness analysis of postoperative goal-directed therapy for high-risk surgical patients. *Crit Care Med* 2014; 42: 1194-203.
49. Pearse RM, Harrison DA, MacDonald N, et al. Effect of a perioperative, cardiac output-guided hemodynamic therapy algorithm on outcomes following major gastrointestinal surgery: a randomized clinical trial and systematic review. *JAMA* 2014; 311: 2181-90.
50. Malgor RD, Alahdab F, Elraiyah TA, et al. A systematic review of treatment of intermittent claudication in the lower extremities. *J Vasc Surg* 2015; 61(3 Suppl): 54S-73S.
51. Moppett IK, Rowlands M, Mannings A, Moran CG, Wiles MD, NOTTS Investigators. LiDCO-based fluid management in patients undergoing hip fracture surgery under spinal anaesthesia: a randomized trial and systematic review. *Br J Anaesth* 2015; 114: 444-59.
52. Jones C, Kelliher L, Dickinson M, et al. Randomized clinical trial on enhanced recovery versus standard care following open liver resection. *Br J Surg* 2013; 100: 1015-24.
53. Kouki P, Matsota P, Christodoulaki K, et al. Greek surgical patients' satisfaction related to perioperative anesthetic services in an academic institute. *Patient Prefer Adherence* 2012; 6: 569-78.

54. Gebremedhn EG, Nagaratnam V. Assessment of patient satisfaction with the preoperative anesthetic evaluation. *Patient Relat Outcome Meas* 2014; 5: 105-10.
55. Moonesinghe SR, Walker EM, Bell M, SNAP-1 Investigator Group. Design and methodology of SNAP-1: a Sprint National Anaesthesia Project to measure patient reported outcome after anaesthesia. *Perioper Med (Lond)* 2015; 4: 4.
56. Black N, Varaganum M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf* 2014; 23: 534-42.
57. Machin JT, Phillips S, Parker M, Carrannante J, Hearth MW. Patient satisfaction with the use of an enhanced recovery programme for primary arthroplasty. *Ann R Coll Surg Engl* 2013; 95: 577-81.
58. Godil SS, Parker SL, Zuckerman SL, et al. Determining the quality and effectiveness of surgical spine care: patient satisfaction is not a valid proxy. *Spine J* 2013; 13: 1006-12.
59. Barnett SF, Alagar RK, Grocott MP, Giannaris S, Dick JR, Moonesinghe SR. Patient-satisfaction measures in anesthesia: qualitative systematic review. *Anesthesiology* 2013; 119: 452-78.
60. Chanthong P, Abrishami A, Wong J, Herrera F, Chung F. Systematic review of questionnaires measuring patient satisfaction in ambulatory anesthesia. *Anesthesiology* 2009; 110: 1061-7.
61. Anonymous. OMERACT, Conference on Outcome Measures in Rheumatoid Arthritis Clinical Trials. Proceedings. Maastricht, The Netherlands, April 29-May 3, 1992. *J Rheumatol* 1993; 20: 527-91.
62. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998; 25: 198-9.
63. Whistance RN, Forsythe RO, McNair AG, et al. A systematic review of outcome reporting in colorectal cancer surgery. *Colorectal Dis* 2013; 15: e548-60.
64. Schmitt J, Apfelbacher C, Spuls PI, et al. The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol* 2015; 135: 24-30.
65. Boers M, Kirwan JR, Gossec L, et al. How to choose core outcome measurement sets for clinical trials: OMERACT 11 approves filter 2.0. *J Rheumatol* 2014; 41: 1025-30.
66. Gargon E, Gurung B, Medley N, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One* 2014; 9: e99111.
67. Royse CF, Newman S, Chung F, et al. Development and feasibility of a scale to assess postoperative recovery: the post-operative quality recovery scale. *Anesthesiology* 2010; 113: 892-905.
68. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011; 343: d5928.
69. Guyatt GH, Oxman AD, Vist G, et al. GRADE guidelines: 4. Rating the quality of evidence-study limitations (risk of bias). *J Clin Epidemiol* 2011; 64: 407-15.
70. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996; 17: 1-12.
71. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med* 2010; 2010(152): 726-32.
72. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010; 8: 336-41.
73. von Elm E, Altman DG, Egger M, et al. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007; 370: 1453-7.
74. Dworkin RH, Turk DC, Farrar JT, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain* 2005; 113: 9-19.
75. Miller TE, Mythen M. Successful recovery after major surgery: moving beyond length of stay. *Perioper Med (Lond)* 2014; 3: 4.
76. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002; 21: 1539-58.