



The need for consensus, consistency, and core outcome sets in perioperative research

Jane M. Blazeby, MD · Paula R. Williamson, PhD · Doug Altman, DSc

Received: 4 June 2015 / Accepted: 4 August 2015 / Published online: 10 December 2015
© Canadian Anesthesiologists' Society 2015

Never before has it been more critical to evaluate the selection and measurement of outcomes. Accurate data, which may be made public, are required for clinical revalidation. As well, institutional level performance data are tied to government goals to achieve targets and make comparisons between centres. Patients require accurate outcome data to be communicated within shared decision-making consultations, and in the research setting, synthesis and meta-analysis of outcomes from well-designed and conducted randomized controlled trials (RCTs) have a critical influence on both health policy and clinical practice. It is vital, therefore, that the relevant stakeholders carefully select and measure the outcomes in these significant evaluations of healthcare interventions in valid and reproducible ways.

In perioperative care and surgical settings, however, clinicians alone have traditionally selected the outcomes for individual performance review, clinical audit and decision-

making, and RCTs. The focus has been mostly on short-term adverse events, and there have been problems with lack of consistency and definition. For example, systematic reviews of bariatric surgery studies have found over 1,000 different outcomes available to evaluate surgery, and over half of these were adverse events.¹ Nevertheless, the outcome definitions in the included papers varied, and no single adverse event or outcome was consistently reported across studies. These problems of multiplicity and heterogeneity are mirrored in perioperative care. A recent Cochrane review of hemodynamic management found that no two studies measured the same outcomes.² The enormous number of different outcomes, which often lack agreed definitions, limit data interpretation and synthesis. This may result in lost opportunities to maximize knowledge from research evidence. It hinders making meaningful comparisons between centres, and it also risks the distortion of research through “outcome reporting bias” where statistically significant findings are more likely to be reported.

Another consistent observation of outcome reporting in perioperative care and surgery studies is the time frame of most outcomes. The majority of papers and most performance reviews for individuals and hospitals focus on data within a 30-day period. In the UK, 30-day mortality rates are publically available for some selected procedures, such as after coronary artery bypass graft surgery. It is uncertain, however, whether this time point is most relevant to patients, as they are rarely consulted when outcomes are selected for routine reporting. In some settings (e.g., major cardiac or cancer surgery or after intensive care admission) 90 days may actually be a more useful time period. Mortality at 90 days may reflect procedure-related complications but is more likely to be caused by recurrence of an underlying problem.

J. M. Blazeby, MD (✉)
Centre for Surgical Research & MRC ConDuCT-II Hub for
Trials Methodology Research, School of Social & Community
Medicine, University of Bristol, Bristol, UK
e-mail: j.m.blazeby@bris.ac.uk

J. M. Blazeby, MD
Division of Surgery, Head & Neck, University Hospitals NHS
Foundation Trust, Bristol, UK

P. R. Williamson, PhD
MRC North West Hub for Trials Methodology Research,
Department of Biostatistics, University of Liverpool, Liverpool,
UK

D. Altman, DSc
Centre for Statistics in Medicine, Nuffield Department of
Orthopaedics, Rheumatology & Musculoskeletal Sciences,
University of Oxford, Oxford, UK

Accordingly, it may be more useful to evaluate provision of care and decision-making. Longer-term outcomes are also important because they can allow for measurement of patient benefit. Benefits are not just the absence of adverse events; they can also reflect symptom relief, improved quality of life, better physical and social function, and a return to normality. In trials assessing clinical effectiveness outcomes, the above patient benefits must be measured along with those more pertinent to healthcare organizations (e.g., resource use and costs) in order to inform how interventions translate into gains for key stakeholders.

One proposed solution to inconsistent selection and measurement of outcomes is to develop “core outcome sets” (COS) and to identify tools for measuring the core outcomes within the set. A COS is an agreed minimum set of health domains (a broad area of health) that should be measured and reported in all studies conducted within a specific clinical area. A COS should include adverse events as well as measurable benefits associated with the intervention. This concept was originally pioneered in cancer research by the World Health Organization (<http://www.comet-initiative.org/studies/details/473?result=true>). The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) initiative (<http://www.omeract.org/>) developed one of the most well-established COS for rheumatoid arthritis. Recent work has shown that, over the last 50 years, an increasing number of RCTs reported using the rheumatoid arthritis COS, and at the same time, there was a decrease in the total number of individual outcomes assessed in those trials.³ This is evidence to show the benefits of a COS and how it can harmonize outcome selection and reporting within a specific condition. There are considerable advantages, therefore, to developing and using a COS for perioperative research, and there are several on-going areas of work in this field.

The Core Outcome Measures in Effectiveness Trials (COMET) initiative supports development of specific methodology and widespread use of COS, including how to involve patients and other stakeholders and how to gain consensus during the COS development (<http://www.comet-initiative.org/>). The COMET also provides a useful online searchable database that summarizes available COS and those under development. The database is underpinned by a large systematic review, updated annually, that identifies all COS publications. The database currently includes over 200 different studies across a variety of areas of health. The COS were developed mostly from the views of clinical and research experts, and some also gained opinion from members of the public.⁴

A multi-professional taskforce, including the European Society of Intensive Care Medicine and the European Society of Anaesthesiology, have undertaken a structured process to establish a list of outcomes to use in

perioperative care studies. This has produced “Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions”, which is a major step towards consistency of outcome reporting in this area.⁵ The standard definitions were derived from a literature search supplemented with papers suggested by an international panel of opinion leaders. The task force appraised the full-text versions of the selected articles, categorized them as relevant (or not), and created four groups of outcomes: adverse events, composite measures, healthcare resource use, and health-related quality of life (HRQL) measures. The outcomes and definitions were selected in a face-to-face meeting. This process identified 22 broad categories of adverse events, four composite measures, and four measures of assessment of resource use and HRQL. There were, however, several weaknesses in the approach used to identify these areas that may limit its validity. For example, there was a lack of patient involvement in the selection of outcomes; in addition, the methods used to reach consensus may have been influenced by characteristics of the participating individuals as opinions and views were not anonymized. It is also possible that the long list of outcomes may not have been drawn from the optimal sources – outcomes important to patients and carers may have been omitted. Despite these areas of concern, it is an important step forward to improve and increase the consistency of outcome selection and reporting in this field.

Parallel and complementary to this work, the Core Outcome Measures for Perioperative and Anaesthetics Care (COMPAC) (<http://www.niaa.org.uk/HSRC-COMPAC>) initiative is developing a COS for perioperative and anesthetic care and is registered in the COMET database (<http://www.comet-initiative.org/studies/details/632?result=true>). The group, led by Professor Mike Grocott, University of Southampton, alongside experts from Monash University, Melbourne, Australia, are using COMET approaches to develop and define outcome measures for widespread use. The group will therefore include patients’ views in the selection of outcomes; hence, it will be interesting to examine how this influences the final core outcome set and compares with the EPCO definitions. This group may generate a COS with the potential to improve and supplement outcome assessment in perioperative care. It will facilitate synthesis of research and improve efficiency. Most importantly, as the core set will be produced in consultation with patients and professionals, it is more likely that the final COS will have outcomes that are highly regarded by both groups of key stakeholders, including longer-term benefits as well as traditional short-term endpoints. When this COS is used in future trials, it is therefore expected to facilitate meaningful decision-making and improve the quality of perioperative research and care.

Le besoin de consensus, de cohérence et d'un ensemble de critères d'évaluation de base en recherche périopératoire

Jamais auparavant n'a-t-il été aussi crucial d'évaluer la sélection et la mesure des critères d'évaluation utilisés dans les études. Des données précises et publiables sont nécessaires pour toute revalidation clinique. Qui plus est, les données de performance au niveau institutionnel sont liées aux objectifs gouvernementaux qui visent à atteindre certaines cibles et établir des comparaisons entre les divers centres. Les patients ont besoin de données d'évaluation précises qui puissent être communiquées lors des consultations de prise de décision partagées. En recherche, la synthèse et la méta-analyse de résultats tirés d'études randomisées contrôlées conçues et réalisées avec rigueur ont une influence décisive sur les politiques de santé aussi bien que sur la pratique clinique. C'est pourquoi il est essentiel que les intervenants pertinents choisissent et mesurent avec soin les résultats de ces évaluations importantes des interventions de santé et ce, de façon valable et reproductible.

Toutefois, lorsqu'il s'agit de soins périopératoires ou de chirurgie, la tradition veut que bien souvent seuls les cliniciens décident des résultats à inclure dans l'évaluation de la performance individuelle, les contrôles et la prise de décision cliniques, et les ERC. L'emphase principale est mise sur les complications à court terme, ce qui a favorisé l'apparition de certains problèmes liés au manque de cohérence et de définition. Par exemple, les revues méthodiques d'études de chirurgie bariatrique rapportent plus de 1000 critères d'évaluation différents utilisés pour évaluer la chirurgie, dont plus de la moitié sont des complications.¹ De plus, les définitions des critères d'évaluation varient entre les articles inclus, et aucune complication ou critère d'évaluation unique n'a été rapporté de façon constante dans toutes les études. Ces problèmes de multiplicité et d'hétérogénéité sont également présents dans les études portant sur les soins périopératoires. Une revue Cochrane récente portant sur la prise en charge de l'hémodynamie a révélé qu'aucune étude ne mesurait les mêmes résultats qu'une autre.² Le nombre gigantesque de critères d'évaluation différents, qui manquent souvent de définitions consensuelles, restreint l'interprétation et la synthèse des données. En raison de cette diversité, nous ratons des occasions de maximiser nos connaissances à partir des données probantes tirées de la recherche. Cela empêche toute comparaison significative

entre les centres, et cela risque de provoquer une distorsion de la recherche en raison d'un « biais de communication des résultats », dans lequel les résultats significatifs d'un point de vue statistique ont plus de chance d'être rapportés.

Une autre constante observée dans la communication des résultats d'études de soins périopératoires et de chirurgie est la période temporelle couverte par la plupart des résultats évalués. La majorité des articles et des rapports de performance individuelle et hospitalière se concentrent sur les données portant sur une période de 30 jours. Au Royaume-Uni, les taux de mortalité à 30 jours sont publics pour certaines interventions, comme par exemple après des pontages aorto-coronariens. Toutefois, nous ne savons pas si cette période est la plus pertinente pour les patients, qui sont rarement consultés lors de la sélection des critères d'évaluation pour la communication habituelle des résultats. Dans certains contextes (par ex., lors de chirurgie cardiaque ou oncologique majeure, ou après une admission aux soins intensifs), une période de 90 jours pourrait en fait être plus utile. La mortalité à 90 jours pourrait refléter les complications liées à l'intervention, mais il est plus probable que le décès soit causé par la récurrence d'un problème sous-jacent. Ainsi, une période de 90 jours pourrait être plus utile pour évaluer les soins et la prise de décision. Les résultats à plus long terme sont également importants parce qu'ils permettent de mesurer les bienfaits aux patients. Ces bienfaits ne se limitent pas uniquement à l'absence de complications; ils peuvent également inclure le soulagement des symptômes, l'amélioration de la qualité de vie, du fonctionnement physique et social, et le retour à la normalité. Dans les études évaluant des critères d'efficacité clinique, les bienfaits pour les patients cités ci-dessus doivent être mesurés simultanément à d'autres résultats plus pertinents aux organismes de santé (par ex., l'utilisation des ressources et les coûts) afin de savoir comment les interventions se traduisent en gains pour les intervenants principaux.

L'une des solutions proposées pour pallier l'hétérogénéité de la sélection et des mesures des résultats est de créer des « ensembles de critères d'évaluation de base » (ou l'acronyme anglais COS, pour '*core outcome set*') et d'identifier des outils permettant de mesurer ces critères d'évaluation de base au sein de l'ensemble. Un COS est un ensemble minimum de domaines de santé (englobant une grande partie de la santé d'un patient) qui devrait être mesuré et rapporté dans toutes les études menées dans un domaine clinique particulier. Un COS devrait inclure les complications ainsi que les bienfaits mesurables associés à une intervention donnée. Ce concept a été utilisé pour la première fois dans la recherche contre le cancer réalisée par l'Organisation mondiale de la Santé (<http://www.>

comet-initiative.org/studies/details/473?result=true). L'initiative OMERACT (*Outcome Measures in Rheumatoid Arthritis Clinical Trials* - <http://www.omeract.org/>) a mis au point l'un des COS les mieux établis à ce jour pour la polyarthrite rhumatoïde. Des travaux récents ont montré qu'au cours des cinquante dernières années, un nombre croissant d'ERC ont été publiées en utilisant le COS pour la polyarthrite rhumatoïde; simultanément, une réduction du nombre total de critères d'évaluation individuels rapportés dans ces études a été observée.³ Il existe des données probantes appuyant les avantages d'un COS et la façon dont il peut harmoniser la sélection des critères d'évaluation et leur communication concernant une maladie en particulier. Ainsi, les avantages liés à la mise au point et à l'utilisation d'un COS sont considérables pour la recherche périopératoire, et plusieurs domaines d'étude sont en cours dans ce domaine.

L'initiative COMET (*Core Outcome Measures in Effectiveness Trials*) appuie la mise au point d'une méthodologie spécifique et l'utilisation répandue de COS, et fournit des pistes concernant notamment la façon d'impliquer les patients et les autres parties intéressées, ainsi que des propositions pour atteindre le consensus au cours de la mise au point d'un COS (<http://www.comet-initiative.org/>). L'initiative COMET met également à disposition une base de données en ligne utile qui résume les COS disponibles et les COS en cours de développement. La base de données est étayée par une importante revue méthodique mise à jour annuellement qui identifie toutes les publications utilisant des COS. À l'heure actuelle, cette base de données compte plus de 200 études différentes dans divers domaines de la santé. Les COS ont été mis au point principalement par les experts cliniques et les chercheurs, mais certains ont également tiré profit de l'opinion de membres du grand public.⁴

Un groupe de travail multiprofessionnel, comptant la Société européenne de réanimation (*European Society of Intensive Care Medicine*) et la Société européenne d'anesthésiologie (*European Society of Anaesthesiology*), a entrepris un processus structurel visant à établir une liste des critères d'évaluation à utiliser dans les études de soins périopératoires. Ce groupe a rédigé des « Normes pour définir et utiliser des mesures d'évaluation pour la recherche sur l'efficacité clinique en médecine périopératoire : les définitions des résultats cliniques périopératoires européennes (EPCO) » (*Standards for definitions and use of outcome measures for clinical effectiveness research in perioperative medicine: European Perioperative Clinical Outcome (EPCO) definitions*). Il s'agit d'une étape importante vers une plus grande cohérence dans la présentation des résultats dans notre domaine.⁵ Les définitions normalisées sont fondées sur une recherche de littérature à laquelle se sont ajoutés

des articles suggérés par un panel international de leaders d'opinion. Le groupe de travail a évalué les versions complètes des articles choisis, les a catégorisés selon leur pertinence et créé quatre groupes de critères d'évaluation : les complications, les mesures composites, l'utilisation des ressources de santé, et les mesures de la qualité de vie liée à la santé (QVLS). Les critères d'évaluation et les définitions ont été retenus dans le cadre d'une réunion en face à face. Ce processus a permis d'identifier 22 catégories générales de complications, quatre mesures composites et quatre mesures d'évaluation de l'utilisation des ressources et de la QVLS. Toutefois, l'approche employée pour identifier ces domaines comportait plusieurs faiblesses, ce qui pourrait nuire à leur validité. Par exemple, on note un manque d'implication de patients dans la sélection des critères d'évaluation; de plus, les méthodes utilisées pour atteindre un consensus ont peut-être été influencées par des caractéristiques propres aux personnes impliquées, étant donné que les opinions et points de vue n'étaient pas anonymes. Il est également possible que la longue liste de critères ne se fonde pas sur les meilleures sources qui soient – certains critères d'évaluation importants pour les patients et les personnes qui en prennent soin auraient pu être omis. Malgré ces inquiétudes, il s'agit d'un progrès important pour améliorer et augmenter la cohérence dans la sélection et la communication des critères d'évaluation dans notre domaine.

En parallèle à ces travaux, et de façon complémentaire, l'initiative COMPAC (*Core Outcome Measures for Perioperative and Anaesthetics Care* - <http://www.niaa.org.uk/HSRC-COMPAC>) est en train de mettre au point un COS pour les soins périopératoires et anesthésiques. Cette initiative est enregistrée dans la base de données COMET (<http://www.comet-initiative.org/studies/details/632?result=true>). Le groupe de travail, sous la direction de Professeur Mike Grocott, de l'Université de Southampton, et avec l'aide d'experts de l'Université Monash de Melbourne en Australie, se sert d'approches de COMET pour mettre au point et définir des critères d'évaluation qui puissent être communément utilisés. Pour ce faire, le groupe va tenir compte de l'opinion de patients dans la sélection de ses critères; il sera donc intéressant de voir dans quelle mesure cette participation influencera l'ensemble final des critères d'évaluation de base, et de le comparer aux définitions de l'EPCO. Ce groupe pourrait créer un COS qui pourrait potentiellement améliorer et s'ajouter à l'évaluation des résultats dans les soins périopératoires. Un tel COS facilitera la synthèse des recherches et améliorera leur efficacité. Mais le plus important à souligner, étant donné que cet ensemble de base sera créé en consultant des patients et des professionnels, c'est que le COS final aura plus de chances d'inclure des critères d'évaluation considérés comme très

importants par ces deux parties, y compris des avantages à plus long terme ainsi que des critères d'évaluation à court terme plus conventionnellement rapportés. Lorsque ce COS sera utilisé dans les études à venir, on peut donc s'attendre à ce qu'il facilite une prise de décision significative et améliore la qualité de la recherche et des soins périopératoires.

Conflicts of interest None declared.

Conflit d'intérêt Aucun.

References

1. Hopkins JC, Howes N, Chalmers K, et al. Outcome reporting in bariatric surgery: an in-depth analysis to inform the development of a core outcome set, the BARIACT Study. *Obes Rev* 2015; 16: 88-106.
2. Gilbert-Kawai ET, Mitchell K, Martin D, Carlisle J, Grocott MP. Permissive hypoxaemia versus normoxaemia for mechanically ventilated critically ill patients. *Cochrane Database Syst Rev* 2014; 5: CD009931.
3. Gargon E, Gurung B, Medley N, et al. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One* 2014; 9: e99111.
4. Kirkham JJ, Boers M, Tugwell P, Clarke M, Williamson PR. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials* 2013; 14: 324.
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.