Article Type: Editorial

EDITORIAL

Satisfaction is not substantially affected by quality of recovery: different constructs or are we lost in statistics?

Satisfaction has been used as a surrogate measure of quality of recovery for many years, principally because it is easy to perform, and is a widely-used feedback tool throughout society. Consider how often people press the 'like' button on various social media platforms! However, the constructs around satisfaction may be very different to the constructs around quality of recovery. This is not to say that satisfaction is not a patient-centered outcome, but rather that it is a different outcome from other quality of recovery domains.

Berning and colleagues investigated the relationship between quality of recovery measured using the QoR-15 scale [1], and compared it with a custom composite satisfaction scale [1]. They concluded that quality of recovery only has a marginal supplementary effect on patient satisfaction. This is a slightly counterintuitive finding, as others have found aspects of quality of recovery such as pain and nausea to influence patient satisfaction [2, 3]. Nonetheless, the article raises the important question of whether patient satisfaction is an adequate measure to assess other quality of recovery outcomes. Although their conclusion pitches this question in the alternative view - that quality of recovery is not an important metric to determine patient satisfaction, outcomes reflecting the quality of recovery, are becoming important in determining whether patients receive value for their surgery. Initiatives such as Enhanced Recovery After Surgery, Peri-operative Medicine, and the Peri-operative Surgical Home are designed to improve the surgical journey, of which the quality of recovery of the patients is an important component. When assessing their finding that there is no meaningful association between quality of recovery and satisfaction,

we need to consider three questions. Firstly, are the constructs of satisfaction and quality of recovery very different and therefore a lack of association is predictable? Secondly, have they used appropriate tools to support their broad conclusion? Thirdly, could their methods of analysis be unsuitable to answer the research question?

Satisfaction and quality of recovery constructs

If the constructs are sufficiently different, then it is unrealistic to expect a close association. If the authors are correct, then satisfaction would be a poor global measure to assess quality of recovery outcomes, and vice-versa. As this field of research and practice develops, it will be important to choose the correct construct to answer the clinical question. If the clinical question revolves around patient-centred outcome quality measures, then satisfaction is unlikely to be a useful measurement. Satisfaction may be heavily influenced by expectation, such that if a patient expects a certain outcome and achieves it, they will be satisfied even if their quality of recovery is poor; whereas a patient who expects a certain outcome and does not achieve it, may be dissatisfied even if their outcome is good. Quality of recovery, however, is a multidimensional construct that is aimed at assessing recovery from the patient's perspective. Quality of recovery outcomes are generally very different from satisfaction outcomes. For example, the ability to return to work is a very different question to "*are you satisfied with your surgical care*"; or "*what is your current pain level*", is a different question to "*are you satisfied with the communication you received from the hospital*".

Tools used to measure satisfaction and quality of recovery

The terms ‘satisfaction’ and ‘quality of recovery’ are ill-defined in the literature [4]. Therefore, we must be careful to interpret the data with reference to the tools used to measure the outcomes. Satisfaction can be an overall question measured with a Likert scale: "*are you satisfied with your care*"; or can be divided into a number of aspects of satisfaction related to expectation, information, comfort, hospital conduct and so forth. Equally, quality of recovery is probably best described as a multidimensional construct, examining physiological, nociceptive, emotive, functional, and cognitive outcomes as well as the overall patient perspective [5]. Further, quality of recovery is not a score or a single time-point, but rather a profile of recovery in each of these domains over time.

There is no perfect tool to measure either satisfaction or quality of recovery. The differences between recovery scales and approaches to assessment of recovery has been summarised by Bowyer et al. [6]. The reader should look carefully at what is being measured when interpreting studies around satisfaction or quality of recovery. For example, the best-known recovery score is the Aldrete score first published in 1970. The Aldrete score [7], is focused on emergence safety and suitability for discharge from the recovery area, and is not suitable as a measure of quality of recovery for subsequent time periods. Wong and colleagues [8] developed the functional recovery index, which measures quality of recovery in three domains; pain and social activity; lower limb activity; and general physical activity. Talamini et al. [9] constructed the Surgical Recovery Index which measures pain with different activities and activity resumption. The concept of DREAMing (drinking, eating and mobilizing) has been proposed as a measurement of anaesthesia care [10]. Myles et al. [11-13] have published three versions of the QoR scale, from a simple nine-point scale to a more comprehensive 40-point scale, using retrospective patient recall over the last 24 hours, and did not include cognition. Royse and colleagues [14] developed the Postoperative Quality of Recovery Scale, which objectively assesses physiological, nociceptive, emotive, functional and cognitive domains and includes overall patient assessment along with satisfaction (using a simple 5-point Likert scale). There are many other quality of recovery or satisfaction scales, and none are capable of fully characterising satisfaction or quality of recovery.

Perhaps the biggest choice in approaches to measuring quality of recovery is whether to collapse the survey items into a single score, or to individualise recovery by comparing each patient to their baseline and dichotomising recovery as 'recovered' or 'not recovered' [6]. The latter approach can preserve recovery information in each individual domain. The major problem with collapsing data into a single score is that one converts a relatively rich data environment into a poor one. Whilst a linear score derived from a series of categorical questions is statistically convenient, it may be clinically meaningless. For example, what does a quality of recovery score of 112 mean? Is it more important for us to identify that, perhaps the main reason for reduced recovery was persistent pain, or depression, or cognitive impairment? Further, it is important to identify whether questions within the survey tools are likely to have a very high ceiling or floor effect, which may also be related to the timing of the survey. For example, if the survey is conducted two weeks after relatively minor surgeryEarnings, then three of the items on the QoR 15 should have a

floor or ceiling effect; those relating to physical comfort, moderate, or severe pain, as it is highly likely that patients will have little or no pain at the time the satisfaction survey is conducted. Equally, satisfaction is likely to change over time, with improved satisfaction occurring when the patient has recovered. Berning et al's [1] conclusion "*quality of recovery has only a marginal supplementary effect on patient satisfaction*" is a statement suggesting high external validity - that it should apply to most types of surgery. Yet, they have conducted a study with specific tools, a single institution, and a limited range of surgical procedures, which would suggest high internal validity, but low external validity. Perhaps their conclusion should be better considered as 'patient satisfaction (using a composite satisfaction scale), measured between one and five weeks after surgery, in a single centre, non-cardiac surgery population, is not associated with the QoR15 scale conducted at four different time-points from baseline to 4 weeks after surgery'.

Statistical approach

The investigators constructed two scales to measure satisfaction and quality of recovery. The questions contained within the satisfaction survey and the QoR-15 are very different, and measured on different scales [11]. For example, in the satisfaction survey there are a range of categorical responses, from a three-point Likert scale to a five or ten point scale, whereas the QoR-15 utilises a visual analogue scale between 0 and 10 for each question [11]. The authors then converted all of these measurement scales into single normalised values between 0 and 100%, and then averaged those measurements to form a total score. For example, in a three-point Likert scale, the worst answer scored 0 and the best answer scores 100%, and we must infer that the middle value scores 50%. The authors then went on to perform correlation analysis between various items of the scales, as well as a multivariable linear regression analysis to identify whether any aspects of the quality of recovery, all satisfaction items (apart from the overall) or baseline characteristics, were correlated with overall satisfaction.

In simplistic terms and in our view, the authors are comparing a variety of fruit (satisfaction) to a variety of vegetables (QoR-15), but by producing a normalised score, they have tried to convert all questions in both surveys to oranges. They then attempted to correlate those two sets of oranges. Such aggregation is usually applied when two sets of questions are essentially believed to be measuring the same underlying latent variable. For

example "*describe your pain as nil, mild, moderate, severe*" vs. "*rate your pain on a scale between 0 and 10*", are clearly expected to be measuring a similar underlying variable. A numerical assessment of the homogeneity of the items combined into these measures is lacking; the authors might have considered, say, the item-total correlation. As noted by Henson [15], consistency in items in a scale is important because it "*speaks directly to the ability of the clinician or the researcher to interpret the composite score as a reflection of the test's items*".

There are also likely to be reliability problems when categorical outcomes of only three levels are used to construct a continuous scale [16]. The analyses used in the study assumes a normal distribution, and accordingly, there are likely violations of the assumptions of normality required for the analysis methods used. Further, most of the satisfaction scale items have a mean value exceeding 90% with many exceeding 95%, which means that the vast majority of patients will have selected the highest category and the items are highly skewed. This also implies a ceiling effect whereby the tool lacks discriminant ability, which reduces the usefulness of the construct [16] . It is difficult to assess the impact of this because, instead of presenting the proportion of patients in each category of the categorical variables, the authors only presented means and standard deviations as summaries of satisfaction. Taking for example, A1 in Table 1, the mean is 97.9 with a standard deviation of 10.0. According to this, the upper limit for a 95% confidence interval would be approximately 117.9 (an impossible value on the 0-100 scale), a clear indication that the assumption of normality used to construct this confidence interval is not valid and, therefore, neither are the regression models used. This choice of presentation also removes the opportunity for the clinician to interpret the data. At no point are we shown the actual data for the study; even the sample sizes are obscured in many cases.

As noted above, problems of skewed data and ceiling effects all reduce the validity when performing linear regression. In each case, categorical variables are treated as continuous, even in the case of type of hospital insurance which seems highly unlikely to be well represented by an interval scale. Also, the model with quality of recovery at time 3 (Model 3) adjusts for demographics and the model with quality of recovery at time 2 (Model 2) does not appear to, which seems hard to justify. It is also not clear how these models were constructed; the methodology implies all ten satisfaction items (excluding overall

satisfaction) were included in the models, but only significant results are presented in the summaries of these models in Table 2.

Implications for future research

When planning future studies around constructs such as quality of recovery or satisfaction, there are number of principles that should, in our view, be considered. Firstly, it is important to define the predominant construct that will be tested: is it satisfaction, quality of recovery, length of stay, or complications? Each of these constructs are different, and one should not use an inappropriate construct as a metric of quality. Secondly, the tools must be appropriate for the task. As a general principle, as quality of recovery is a multidimensional construct, then one should resist the temptation to collapse all those domains and produce a single outcome score. Whilst this is statistically convenient, it may be clinically meaningless. It is important to consider whether the patient should be assessed at baseline to determine their pre-surgery functional and cognitive status, and then postoperative outcomes can be compared to their pre-operative status. This allows investigators to individualise quality outcomes, rather than to collapse multiple individual differences into group data. We also believe that the quality of recovery tool chosen should be broad enough to measure what is intended, and the conclusion must be limited to the discriminant ability of the tool. For example, one should not comment on cognitive recovery if the quality of recovery tool does not measure it. The term quality of recovery is so broad, that it should be defined more precisely according to what is measured when analysing the results and making conclusions. In terms of statistical analysis, a more common approach to assessing an outcome which suffers from ceiling effects, is to dicotomise the outcome (for example satisfaction) and analyse this using a multivariable logistic regression method, where the explanatory variables are included as categorical or continuous as appropriate. This will produce predictors of satisfaction, with odds ratios and confidence intervals. We also suggest that the actual frequency data for the measurements always be available to the reader, in order to have a better understanding of the clinical relevance of the findings.

Conclusion

The measurement of patient-reported outcomes after surgery is a relatively new field of audit and research in anaesthesia practice. Satisfaction and quality of recovery are both

patient-reported outcomes, but with different constructs, and therefore are different outcomes. It is important when evaluating the literature to identify what is actually being measured, as at this stage of development, both satisfaction and quality of recovery are dependent on the survey tools used to measure the outcomes. We suggest that the temptation to generalise the findings of specific survey tools to the more global and poorly-defined concepts of satisfaction and quality of recovery should be resisted. We also suggest that care must be taken to avoid using satisfaction as a surrogate measure of quality of recovery and vice-versa.

Acknowledgements

CR is the co-originator for the Postoperative Quality of Recovery Scale. No external funding or other competing interests declared.

C. F. Royse

Professor, Department of Surgery, The University of Melbourne and
Consultant Anaesthetist, Royal Melbourne Hospital
Email: Colin.royse@unimelb.edu.au

S. Clarke

Statistical Consultant, Statistical Consulting Centre, The University of Melbourne, Melbourne, Australia

Keywords: quality measures, patient care; recovery, postoperative; patient satisfaction
Accepted: 6 April 2017

Typesetter, please add a footnote on the bottom left of the 1st page saying:

This editorial accompanies an article by Berning et al., Anaesthesia 2017; 72: doi: 10.1111/anae.xxxxx.

References

1. Berning V, Laupheimer M, M N, Heidegger T. Influence of quality of recovery on patient satisfaction with anaesthesia and surgery: a prospective observational cohort study. *Anaesthesia* 2017; 72: doi: 10.1111/anae.xxxxx.

2. Royse CF, Chung F, Newman S, Stygall J, Wilkinson DJ. Predictors of patient satisfaction with anaesthesia and surgery care: a cohort study using the Postoperative Quality of Recovery Scale. *European Journal of Anaesthesiology* 2013; **30:** 106-10.

3. Myles PS, Williams DL, Hendrata M, Anderson H, Weeks AM. Patient satisfaction after anaesthesia and surgery: results of a prospective survey of 10,811 patients. *British Journal of Anaesthesia* 2000; **84:** 6-10.

4. Bowyer AJ, Royse CF. Postoperative recovery and outcomes - what are we measuring and for whom? *Anaesthesia* 2016; **71 Suppl 1:** 72-7.

5. Bowyer A, Royse C. The importance of postoperative quality of recovery: influences, assessment, and clinical and prognostic implications. *Canadian Journal of Anesthesia* 2016; **63**: 176-83.

6. Bowyer A, Jakobsson J, Ljungqvist O, Royse C. A review of the scope and measurement of postoperative quality of recovery. *Anaesthesia* 2014; **69:** 1266-78.

7. Aldrete JA, Kroulik D. A postanesthetic recovery score. *Anesthesia and Analgesia* 1970; **49:** 924-34.

8. Wong J, Tong D, De Silva Y, Abrishami A, Chung F. Development of the functional recovery index for ambulatory surgery and anesthesia. *Anesthesiology* 2009; **110:** 596-602.

9. Talamini MA, Stanfield CL, Chang DC, Wu AW The Surgical Recovery Index. *Surgical Endosccopy* 2004; **18:** 596-600.

10. Levy N, Mills P, Mythen M. Is the pursuit of DREAMing (drinking, eating and mobilising) the ultimate goal of anaesthesia? *Anaesthesia* 2016; **71:** 1008-12.

11. 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.

12. Myles PS, Weitkamp B, Jones K, Melick J, Hensen S. Validity and reliability of a postoperative quality of recovery score: the QoR-40. *British Journal of Anaesthesia* 2000; **84:** 11-5.

13. Myles PS, Hunt JO, Nightingale CE, et al. Development and psychometric testing of a quality of recovery score after general anesthesia and surgery in adults. *Anesthesia and Analgesia* 1999; **88:** 83-90.

14. 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.

15. Henson RK. Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development* 2001; **34:** 177-89.

16. Steiner D, Norman G, Caimey J. *Health measurement scales: a practical guide to their development and use.* USA, Oxford University Press, 2014.