Author/s:
Bowyer, AJ;Heiberg, J;Sessler, DI;Newman, S;Royse, AG;Royse, CF

Title:
Validation of the cognitive recovery assessments with the Postoperative Quality of Recovery Scale in patients with low-baseline cognition

Date:
2018-11-01

Citation:
Bowyer, A. J., Heiberg, J., Sessler, D. I., Newman, S., Royse, A. G. & Royse, C. F. (2018). Validation of the cognitive recovery assessments with the Postoperative Quality of Recovery Scale in patients with low-baseline cognition. Anaesthesia, 73 (11), pp.1382-1391. https://doi.org/10.1111/anae.14402.

Article Type: Original Article

ORIGINAL ARTICLE

Validation of the cognitive recovery assessments with the Postoperative Quality of Recovery Scale in patients with low baseline cognition

A. J. Bowyer [1], J. Heiberg [2], D. I. Sessler [3], S. Newman [4], A. G. Royse [5], C. F. Royse [6]

1. *Specialist Anaesthetist, The Royal Melbourne Hospital, Melbourne, Australia. andrea.bowyer@hotmail.com*

2. Associate Professor, Dept. of Cardiothoracic & Vascular Surgery, Aarhus University Hospital and Dept. of Clinical Medicine, Aarhus University, Denmark. johan.heiberg@clin.au.dk

3. Michael Cudahy Professor & Chair, Department of ***O****UTCOMES* ***R****ESEARCH, Anesthesiology Institute, Cleveland Clinic, Cleveland, USA. ds@or.org*

4. Pro- Vice Chancellor, City University London, London, United Kingdom. Stanton.Newman.1@city.ac.uk

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/anae.14402

5. Professor of Surgery, The University of Melbourne, Melbourne, Australia.
Alistair.royse@unimelb.edu.au

6. Professor of Anaesthesia, The University of Melbourne, Melbourne, Australia.
Colin.royse@unimelb.edu.au

Correspondence to: Colin Royse

Department of Surgery, The University of Melbourne

Level 6, Centre for Medical Research, Royal Melbourne Hospital

Parkville VIC 3050

E: colin.royse@unimelb.edu.au

Short running title: PostopQRS low-baseline cognition recovery validation

Keywords: anaesthesia; POCD; ageing: CNS changes; cardiac surgery; cognitive recovery

Accepted: 6 June 2018

Summary

Patients with pre-surgery cognitive impairment cannot currently be assessed for cognitive recovery after surgery using the Postoperative Quality of Recovery Scale (PostopQRS), as they would mathematically be scored as recovered. We aimed to validate a novel method to score cognitive recovery in patients with low baseline cognition, using the number of low-score tests rather than their numerical values. Face validity was demonstrated in 86 participants in whom both the Postoperative Quality of Recovery Scale and an 11-item neuropsychological battery were performed. The Postoperative Quality of Recovery Scale agreed with neuropsychological categorization of low versus normal cognition 74% of the time, with all but 5 incorrectly coded participants deviating by only 1 neurocognitive test. Cognitive recovery over time was comparable for groups with differing baseline cognitive function, irrespective of whether the Postoperative Quality of Recovery Scale or neuropsychological methods were used. Discriminant validation was demonstrated in a post hoc analysis of the Steroids in Cardiac Surgery (SIRS) sub study by allocating groups to normal (n=246) or low baseline cognition (n=231) stratified by cognitive recovery on day 1. Recovery was similar for participants with low and normal baseline cognition. Postoperative length of stay was longer in patients with failed cognitive recovery whether they had normal (10.4 ± 10.0 vs. 8.0 ± 5.9 days, P=0.02) or low baseline cognition (12.0 ± 11.1 vs. 8.2 ± 4.7 days, P<0.01). Overall quality, as well as cognitive, emotive, and physiological recovery was independent of baseline cognition. The modified scoring method for the Postoperative Quality of Recovery Scale cognitive domain demonstrates acceptable face and discriminant validity.

Post-operative cognitive dysfunction (POCD) is associated with serious morbidity and mortality [1-5]. Cognitive decline is common after major surgery [6-9], particularly in the elderly [3,8], making cognitive recovery an important determinant of overall long-term surgical outcome. Early detection of cognitive recovery may alert health care practitioners to the risk of poor overall recovery and, possibly, provide opportunity for early intervention. The Postoperative Quality of Recovery Scale (PostopQRS) is a multidimensional survey-based tool, which includes a domain designed to measure cognitive postoperative recovery

over time [10]. The cognitive domain includes five verbal cognitive tests. The other domains are physiological, emotive, nociceptive, and activities of daily living. The Scale also includes an overall patient perspective domain.

The scale is determined preoperatively to provide individual baseline measurements for each patient. Thereafter, recovery in various domains is defined by postoperative values equaling or exceeding individual baseline values, except for the cognitive domain where a tolerance level is added to adjust for normal performance variability [11]. That is, patients can perform a little worse than baseline in the cognitive domain, by the magnitude of the tolerance factor, and still be scored as recovered. Currently, patients whose preoperative performance is so low such that any postoperative performance will fall within the tolerance factor, are considered to have a low baseline performance in that test. A consequence of having low baseline cognition on a test preoperatively is that these patients will be automatically scored as "recovered" when assessed postoperatively for these tests. Accordingly, participants with low baseline cognition are not currently scored in the cognitive domain postoperatively, although they can be scored in other recovery domains. The difficulty is that patients with pre-existing cognitive impairment are at especially great risk of poor outcomes, especially if their cognition deteriorates [1,2,12]. To address this limitation of the PostopQRS, we developed a novel scoring method for cognitive recovery in patients with low baseline scores in the cognitive domain. Specifically, we now consider the number of cognitive tests (out of 5) on which a patient performs poorly preoperatively (i.e. below the test's threshold value for low baseline cognition). Postoperative recovery in this population is then defined as the same number of tests or fewer that score below the threshold value. If more tests score below the threshold values than seen preoperatively, then cognitive recovery has not occurred. Our aim was to demonstrate both face and discriminant validity for this modified scoring method.

Methods

We used two existing datasets. The first was from a single-centre observational study entitled "Comparison of neurocognitive assessment vs. PostopQRS cognitive domain performance to assess cognitive recovery in patients undergoing cardiac surgery", conducted at the Royal Melbourne Hospital, Australia (Melbourne Health HREC 2011.142). It included 69 patients undergoing cardiac surgery, each of whom had both PostopQRS and a comprehensive neurocognitive test battery. The second was from the Quality of Recovery sub-study of the Steroids In caRdiac Surgery (SIRS) trial [13]. All participants understood and spoke English well enough to complete the testing surveys, and none had known psychiatric disease, dementia, or any medical or learning disorder that would impair cognitive ability.

In each study, patients were assessed preoperatively and at multiple postoperative times. We compared cognitive recovery in patients who had normal or low cognitive baseline and determined whether clinical outcomes (quality of patient recovery, length of stay, cardiovascular complications, surgical complications and death) differed for participants who did and did not demonstrate cognitive recovery on the first postoperative day.

The cognitive domain of the PostopQRS consists of 5 verbal cognitive tests: orientation, digits forward, digits back, word recall, and word generation. Cognition is deemed recovered when scores on all five tests return to within a small delta ('tolerance factor') of baseline values. The tolerance factors for the cognitive subdomains are 0/**3** for orientation, 2/**6** for digits forward, 1/**6** for digits back and 3/**15** for word recall, and 3/unlimited for word generation [11]. Since patients whose initial cognitive performance on any test is less than the allowed tolerance are automatically scored as recovered postoperatively with the original scoring approach, we modified our scoring system for patients whose preoperative cognitive function was poor, with poor being defined by an initial score within the tolerance range on any of the 5 cognitive tests. Specifically, the modified system considers the number of cognitive tests on which patients scores below the tolerance range preoperatively. Patients are then considered recovered postoperatively when they score poorly on no more tests than they did preoperatively. For example, a patient who had 1 low baseline cognitive test preoperatively will be scored as recovered if they perform poorly on

no more than any 1 of the 5 cognitive tests postoperatively, but not if they have low performance on 2 or more of the tests.

Face validation: ***Comparison of neurocognitive assessment vs. PostopQRS cognitive domain performance to assess cognitive recovery in patients undergoing cardiac surgery.***

We determined face validity by comparing low versus normal baseline patients using an 11-item neurocognitive battery as the reference method. Face validity was considered to have been established if cognitive recovery was similar for low and normal baseline cohorts using each method. The PostopQRS and the neuropsychological battery tests were performed by trained research staff. The PostopQRS interviews were conducted face-to-face while patients were hospitalized or via telephone after discharge (11). The neurocognitive testing coincided with the corresponding face-to-face PostopQRS interviews, with both tests being performed by the same research team member. Baseline PostopQRS and neurocognitive baseline data collection occurred no more than 14 days prior to surgery. Subsequent PostopQRS data collection occurred on postoperative day 1, days 3-5, day 14, day 30, and weeks 6-8. Postoperative neurocognitive testing occurred at postoperative days 3-5 and weeks 6-8. Each test was carried out by a trained research member, with blinding occurring at data entry and analysis.

The neuropsychological battery included 11 validated and widely utilised neurocognitive tests that assessed domains recommended in the Consensus Statement on postoperative neurobehavioral outcomes [14,15].Test outcomes were either time to test completion (Trails A & B and Grooved Pegboard) or number of correct responses (Controlled Oral Word Association Test, Stroop Colour Word Test, Rey Auditory Verbal Learning Test, Longest Digit Span (Forward and Backward) and Symbol Digit Modality). Where appropriate, reference parameters of these tests were age and sex matched and parallel forms containing different word and number lists were used in order to reduce learning effect [16-18].

The analysis of the neuropsychological tests was conducted as follows.

1. A literature search was conducted to identify age- and sex-matched reference values for each test (mean and standard deviation), which were used as reference values to standardize each participant's test performance and are shown in Table S1.

2. Baseline neuropsychological performance was determined by comparison of the participant's preoperative neuropsychological test scores to that of the age- and sex-matched population norms [16-18], expressed as Z scores (standard deviations from the mean). Baseline neuropsychological cognition was considered poor when patients had a normalized Z scores less than -1.96 in at least two of the 11 tests in the battery which is consistent with PostopQRS scoring in patients were categorized as having low baseline cognition when they had a low baseline score in at least one of the 5 tests in the battery, as well as current definitions of poor neuropsychological performance

The number of tests in which a patient achieved normal baseline scores was compared for both neuropsychological and PostopQRS tests. This process quantified agreement between the reference neuropsychological method and the PostopQRS in categorizing low versus normal baseline participants. Cognitive recovery in the PostopQRS domain was calculated using the current scoring method for normal baseline participants, and the modified scoring method for low-baseline participants. PostopQRS cognitive recovery, as defined by both PostopQRS and neurocognitive testing, was compared in normal and low-baseline participants using Cochran–Mantel–Haenszel tests. We were, thus, able to determine whether cognitive recovery, as measured by the PostopQRS, was similar amongst groups classified as having low versus normal baseline cognition. A reverse analysis was also performed. Neuropsychological recovery was graphed for days 3-5 and weeks 6-8 for both low and normal baseline groups, with baseline group classification being determined both by the PostopQRS and neuropsychological definitions. This approach allowed us to determine whether neuropsychological recovery was similarly defined by PostopQRS and neuropsychological definitions of low versus normal baseline cognition.

Discriminant validation: ***The impact of early cognitive recovery after cardiac surgery: an analysis of the SIRS trial sub study [13]***

To investigate discriminant validation, we hypothesised that quality-of-recovery as well as clinical outcomes (length of stay, cardiovascular complications, surgical complications and death) would be worse in patients who failed to recover early (day 1) in the cognitive domain, compared to those patients who did have early cognitive recovery, independent of whether they had low or normal baseline cognitive function. This sub-study was part of the Steroids In SIRS trial (ClinicalTrials.gov identifier NCT00427388).[19] The sub-study started well after the underlying trial began, and was restricted to the Royal Melbourne Hospital (Australia), the Cleveland Clinic (USA), and the Hamilton Health Sciences and McMaster University (Canada). Amendments to the SIRS ethics approvals and local governance approvals were obtained at each participating centre to conduct this sub study.

A detailed description of the underlying SIRS trial has been published [19]. In brief, it was a double-blind, placebo-controlled, randomised, multicentre trial on high-risk patients undergoing cardiac surgery (EuroSCORE ≥6). The intervention was to administer 250 mg of methylprednisolone at anaesthetic induction and a further 250 mg just prior to commencement of cardiopulmonary bypass. The PostopQRS was conducted within 2 weeks before surgery (baseline) and 1 day, 2 days, 3 days, 1 month, and 6 months after surgery. The physiological subdomain was only measured on postoperative days 1 to 3. The results of the comparison between methylprednisolone and placebo have been previously reported and showed no difference between groups [20]. In this analysis, the participants were, therefore, combined.

Cognitive recovery was scored using the current method for participants with normal baseline scores and using the modified scoring system for participants with low baseline scores (see above). Two groups were defined *post hoc* according to the presence or absence of cognitive recovery on day 1 after surgery. The primary outcome used to determine discriminant validation was length of hospital stay after cardiac surgery. Secondary outcomes used to support discriminant validation included other clinical adverse events, including death, atrial fibrillation, myocardial infarction, stroke, acute kidney injury, and infections within 6 months after cardiac surgery as well as failure to recover in the overall

domain and the physiological, nociceptive, emotive, functional, and cognitive subdomains over 6 months following cardiac surgery.

Continuous data are presented as means ± standard deviations, binary data are presented as absolute numbers and percentages of participants, and odds ratios are displayed with 95% confidence intervals (CI). Differences in clinical events were assessed using students *t*-tests for continuous data and chi-squared tests for binary data. In terms of quality of recovery over time, differences between groups were assessed using the Cochrane-Mantel-Haenszel test. In case of a significant difference between groups in quality of recovery over time, chi-squared tests were used to compare the proportions of participants recovered at each individual time point. Statistical analysis was not performed for the evaluation of severity of incomplete cognitive recovery due to an inadequate sample size to exclude Type II error. P-values < 0.05 were considered statistically significant. All *p*-values are two-sided. Descriptive data were stored in Microsoft Excel 2016 (Microsoft Corp., CA, USA), statistical analyses were performed using Stata/IC 12.1 for Mac (Stata Corp., TX, USA), and data were graphically described in Graph Pad Prism 6 (GraphPad Software, La Jolla, CA, USA). A sample size was determined by the number of participants assessed with the PostopQRS in the SIRS studies, therefore no formal sample size estimates were performed.

Results

Face validation: ***Comparison of neurocognitive assessment vs. PostopQRS cognitive domain performance to assess cognitive recovery in patients undergoing cardiac surgery***

Sixty-nine participants were enrolled, but two were excluded because of incomplete data. Eighteen participants (27%) had low baseline PostopQRS cognition. Demographic and operative data are shown in Table S2. For each participant, the number of tests that were scored as normal for both PostopQRS and for neuropsychological batteries are shown in

Figure 1. There was agreement in categorisation in 74% of participants. Where categorization differed, the magnitude of difference was 1 test on the neuropsychological battery in all but 5 participants. Among the 11 patients who had a normal PostopQRS but low neuropsychological at baseline, 9 failed in at least one motor test. Of those with normal neuropsychological baseline but low PostopQRS cognition, 4/6 failed one neuropsychological verbal test. Cognitive recovery in the PostopQRS domain is shown for normal and low baseline participants in Figure 2. Cognitive recovery was similar for normal and low baseline participants, irrespective of the definition used to categorise low baseline participants. Neuropsychological recovery on days 3-4 and at 6-8 weeks is shown in Figure 3. Recovery in the normal baseline group was similar to that in the low baseline group, irrespective of the definition used to categorise low baseline participants.

Discriminant validation: ***The impact of early cognitive recovery after cardiac: an analysis of the SIRS trial sub study [13]***

From January 2012 to December 2013, 555 participants were enrolled of whom 48 had incomplete baseline scores and 27 had missing follow-up data on the first day after surgery. There were, thus, 482 participants available for data analyses (246 normal cognitive baseline and 234 low cognitive baseline, Figure S1). In participants with normal baseline cognitive scores 79/246 participants, and for low baseline cognition 89/234 participants had cognitive recovery on the first postoperative day.

Baseline characteristics and operative details for normal and low baseline cognition groups are shown in Table S3. Mean ages were 72 $\pm$ 12 years in the cognitively recovered group, and 72 $\pm$ 11 years in the non-recovered group. Generally, the cognitively recovered and non-recovered participants were similar, although there was a tendency towards a higher proportion of active smokers among the non-recovered participants. The normal and the low baseline groups were similar except the tendency towards higher age in the low baseline group.

A comparison of recovery in all domains of the PostopQRS, using the modified scoring system for low baseline cognition, is shown for all participants with normal and low baseline cognition in Figure S2. The low baseline group had slightly better cognitive recovery, but with a similar profile across time to the patients with normal baseline cognition. Recovery profiles did not differ significantly in any other domains.

Clinical outcomes of participants with normal cognitive baseline scores are displayed in Table 1. The total length of hospitalization in the cognitively recovered group was 8.0 $\pm$ 5.9 vs. 10.4 $\pm$ 10.0 days, p=0.02). The proportion of participants with a length of stay beyond 10 days was 13% in the recovered group versus 25% in the non-recovered group, (odds ratio: 2.4, 95% CI 1.1-5.0, p=0.02). Length-of-stay in the intensive care unit (ICU) was 2.1 $\pm$ 3.0 days versus 3.2 $\pm$ 5.5 days, p=0.049; and the proportion of participants with a length of ICU stay beyond 2 days was 14% in the recovered group, which was lower than 34% in the non-recovered group, (odds ratio: 3.1, 95% CI 1.5-6.3, p=0.01). There were no differences between groups in any of the other outcomes.

Quality of recovery over time to 6 months after surgery in normal baseline cognition participants is shown for cognitively recovered versus not recovered groups in Figure 4. The incidence of overall recovery was higher for recovered compared to non-recovered participants, p<0.01. The cognitive and physiological domains showed better recovery for the recovered group (p<0.01), but not emotive, nociceptive or functional recovery.

The clinical outcomes for participants with low baseline cognitive scores are shown in Table 2. The total length of stay was shorter in the cognitively recovered group (8.2 $\pm$ 4.7 vs. 12.0 $\pm$ 11.1 days, p<0.01), as was the proportion of participants with a length of stay beyond 10 days (16% vs. 37%, p<0.01, odds ratio: 3.2, 95% CI 1.7-6.2). Length of ICU stay was shorter in the cognitively recovered group (2.4 $\pm$ 2.5 vs. 3.9 $\pm$ 4.7, p<0.01), as was the proportion of participants with a length of ICU stay beyond 2 days (19% vs. 42%, p<0.01, odds ratio: 3.1, 95% CI 1.7-5.8). In addition, the incidence of death at 30 days after randomization was lower in cognitively recovered participants (1% vs. 9%, p=0.01). The

groups were generally similar in terms of the other adverse outcomes. Quality of recovery over time for patients with low cognitive baseline scores is illustrated in Fig. 5. Overall recovery was higher in the recovered group (p<0.01). Cognitive, physiological and emotive recovery was higher in the recovered group (all p<0.01), whereas there were no differences between the groups in terms of nociception or activity of daily living.

The recovery profiles for participants with normal cognitive baseline scores based on the number of tests where recovery had failed, are shown in Figure S3 and participants with low cognitive baseline scores are displayed in Figure S4. Generally, recovery was worse in patients with poor baseline cognitive function, especially in the cognitive, physiological, and overall recovery domains.

Discussion

Face validity implies that the test responds as expected and includes construct validity where tests measuring similar constructs should produce similar results [21]. The PostopQRS agreed with neuropsychological categorisation of low versus normal baseline cognition in most participants. Importantly, where disagreement occurred, it was usually by a small margin on 1 test on the neuropsychological battery. This is unsurprising, as PostopQRS cognitive tests are derivatives of the neuropsychological battery, with a focus on brief, verbal tests designed to facilitate telephone use [11]. The 11–item battery includes visuospatial and motor tests in addition to the verbal tests, and though they essentially measure similar constructs, the motor tests assess cognitive performance in areas that the PostopQRS does not.

Most participants with normal baseline PostopQRS but low baseline neuropsychological cognition failed at least one motor test, whereas the majority of low baseline PostopQRS/normal baseline neuropsychological participants failed a single verbal test. It is of note that motor tests are not included in the PostopQRS Cognitive subtests. We demonstrated further face validity by showing that cognitive recovery was similar for both

low and normal baseline cohorts, whether or not PostopQRS or neurocognitive methods were used to categorise low baseline cognition. We included a reverse analysis, and showed face validity that a neurocognitive definition of recovery was comparable in normal- and low-baseline participants, irrespective of the method used to categorize low baseline.

Recovery improved over time in both normal and low baseline groups in a manner that is consistent with previous data [10,11,22-25], indicating that the new scoring system produces similar results in low baseline participants to the normal baseline participants. This similarity of cognitive recovery between groups that have normal vs. low baseline cognition has potentially important clinical implications given the current debate regarding the relationship between pre-existing cognitive impairment and poor postoperative cognitive function [1,3,26-29], and emphasises the need to validate a scoring system with which to measure cognitive recovery in those patients with a low preoperative baseline.

We demonstrated discriminant validity for the modified scoring system using the SIRS trial sub study data [13]. Quality of recovery in the SIRS trial was similar in normal and low baseline participants for domains other than cognition (which could previously not be scored for the low baseline group). However, when the modified cognitive scoring was applied to the low baseline participants, and groups were allocated according to cognitive recovery on day 1, then the differences in clinical and quality of recovery outcomes in both low and normal baseline groups were similar, favouring early cognitive recovery. Discriminant validity was further demonstrated by a pattern of increasingly worse recovery as the number of tests failed at baseline increased, indicating a severity effect.

Our primary aim was to validate a new scoring method for cognitive recovery in patients with low baseline cognition. The cohort study comparing PostopQRS and the neuropsychological test was sufficient to demonstrate face validity, but too small to evaluate clinical outcomes. In the clinical trial, we allocated groups according to cognitive recovery on day 1 and, hence, only participants who completed the PostopQRS at Day 1 were included in these analyses. It is possible that participants who declined to complete the assessment at this time point had a worse outcome introducing a potential inclusion

bias. The sample size in this sub study was determined by the number of participants assessed with the PostopQRS in the SIRS study, and it was not *a priori* powered for the post hoc analysis to detect potential differences between low and normal cognitive baseline scores. Comparison of these groups was used to test for discriminant validation of the proposed scoring system for low baseline patients, rather than to assess the importance of early cognitive recovery, and we urge caution in further analysis of these data. However, whether early detection of cognitive failure provides a window-of-opportunity for interventions that might improve outcomes, merits further research.

Acknowledgments

The authors thank Mr Matt Lanchaster and Ms Zelda Williams for assistance with the conduct of the face validation study.

No funding was obtained to conduct the studies. The PostopQRS has previously received funding from Baxter HealthCare, but not specifically for this project. Colin Royse and Stanton Newman helped develop the PostopQRS, and are directors of PostopQRS Limited.

Supporting information

Table S1. Age and gender matched values for the neuropsychological tests. LDSF – Longest Digit Span Forward; LDSB – Longest Digital Span Backward; Trail A & B – Trail Tests A & B; COWAT – Controlled Oral Word Association Test; SCWT – Stroop Colour Word Test – Interference (Trennery); GP-Dom – Grooved Pegboard, dominant hand; GP-NonDom - Grooved Pegboard, non-dominant hand; RAVLT 1-5 – Rey Auditory Verbal Learning Test, total (tests 1-5); RAVLT – Recall - Rey Auditory Verbal Learning Test, recall; SDMT – Symbol Digit Modality Test.

Table S2. Demographic and operative details for the face validation study "Comparison of neurocognitive assessment vs. PostopQRS cognitive domain performance to assess cognitive recovery in patients undergoing cardiac surgery."

Table S3. Characteristics and operative details of cardiac surgery patients (SIRS sub study) with either normal or low cognitive baseline scores.

Figure S1. Participant flow diagram for the SIRS sub study.

Figure S2. Recovery over time is shown for the SIRS sub study cohort categorized by normal or low baseline cognition.

Figure S3. Recovery over time is shown for the SIRS sub study cohort with normal baseline cognition, and categorized according to the number of tests that the participants failed recovery.

Figure S4. Recovery over time is shown for the SIRS sub study cohort with *low baseline cognition*, and categorized according to the number of tests that the participants failed recovery.

References

1. Monk TG, Weldon BC, Garvan CW, et al. Predictors of cognitive dysfunction after major noncardiac surgery. *Anesthesiology* 2008; **108:** 18-30.
2. Price CC, Garvan CW, Monk TG Type and severity of cognitive decline in older adults after noncardiac surgery. *Anesthesiology* 2008; **108:** 8-17.
3. Moller JT, Cluitmans P, Rasmussen LS, et al. Long-term postoperative cognitive dysfunction in the elderly ISPOCD1 study. ISPOCD investigators. International Study of Post-Operative Cognitive Dysfunction. *Lancet* 1998; **351:** 857-61.

4. Steinmetz J, Christensen KB, Lund T, Lohse N, Rasmussen LS, Group I Long-term consequences of postoperative cognitive dysfunction. *Anesthesiology* 2009; **110:** 548-55.
5. Phillips-Bute B, Mathew JP, Blumenthal JA, et al. Association of neurocognitive function and quality of life 1 year after coronary artery bypass graft (CABG) surgery. *Psychosom Med* 2006; **68:** 369-75.
6. Rasmussen LS, Steinmetz J Ambulatory anaesthesia and cognitive dysfunction. *Current opinion in anaesthesiology* 2015; **28:** 631-5.
7. Rasmussen LS, Larsen K, Houx P, Skovgaard LT, Hanning CD, Moller JT The assessment of postoperative cognitive function. *Acta anaesthesiologica Scandinavica* 2001; **45:** 275-89.
8. Canet J, Raeder J, Rasmussen LS, et al. Cognitive dysfunction after minor surgery in the elderly. *Acta anaesthesiologica Scandinavica* 2003; **47:** 1204-10.
9. van Dijk D, Keizer AM, Diephuis JC, Durand C, Vos LJ, Hijman R Neurocognitive dysfunction after coronary artery bypass surgery: a systematic review. *The Journal of thoracic and cardiovascular surgery* 2000; **120:** 632-9.
10. 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.
11. Royse CF, Newman S, Williams Z, Wilkinson DJ A human volunteer study to identify variability in performance in the cognitive domain of the postoperative quality of recovery scale. *Anesthesiology* 2013; **119:** 576-81.
12. Steinmetz J, Siersma V, Kessing LV, Rasmussen LS Is postoperative cognitive dysfunction a risk factor for dementia? A cohort follow-up study. *British journal of anaesthesia* 2013; **110 Suppl 1:** i92-7.
13. Royse CF, Saager L, Whitlock R, et al. Impact of Methylprednisolone on Postoperative Quality of Recovery and Delirium in the Steroids in Cardiac Surgery Trial: A Randomized, Double-blind, Placebo-controlled Substudy. *Anesthesiology* 2017; **126:** 223-33.
14. Murkin JM, Newman SP, Stump DA, Blumenthal JA Statement of consensus on assessment of neurobehavioral outcomes after cardiac surgery. *The Annals of thoracic surgery* 1995; **59:** 1289-95.
15. Stump DA Selection and clinical significance of neuropsychologic tests. *The Annals of thoracic surgery* 1995; **59:** 1340-4.
16. Strauss E, Sherman EMS, Spreen O. *Compendium of neuropsychological tests A compendium of neuropsychological tests : administration, norms, and commentary (3rd ed).* New York, Oxford University Press, 2006.

17. Wechsler D. *Wechsler Adult Intelligence Scale - Fourth Edition Australian and New Zealand Language Adapted Edition (WAIS-IV A&NZ Language Adapted Edition)* Sydney, Australia, Pearson Clinical and Talent Assessment, 2008.
18. Trennery M, Crosson B, Deboe J, Leber W. *Stroop Neurospychological Screening Test Manual* Florida, PAL, 1989.
19. Whitlock RP, Devereaux PJ, Teoh KH, et al. Methylprednisolone in patients undergoing cardiopulmonary bypass (SIRS): a randomised, double-blind, placebo-controlled trial. *Lancet* 2015; **386:** 1243-53.
20. Royse CF, Saager L, Whitlock R, et al. Impact of Methylprednisolone on Postoperative Quality of Recovery and Delirium in the Steroids in Cardiac Surgery Trial: A Randomized, Double-blind, Placebo-controlled Substudy. *Anesthesiology* 2016.
21. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology* 2007; **60:** 34-42.
22. Lindqvist M, Schening A, Granstrom A, Bjorne H, Jakobsson JG Cognitive recovery after ambulatory anaesthesia based on desflurane or propofol: a prospective randomised study. *Acta anaesthesiologica Scandinavica* 2014; **58:** 1111-20.
23. Royse CF, Williams Z, Ye G, et al. Knee surgery recovery: Post-operative Quality of Recovery Scale comparison of age and complexity of surgery. *Acta anaesthesiologica Scandinavica* 2014; **58:** 660-7.
24. Royse CF, Williams Z, Purser S, Newman S Recovery after nasal surgery vs. tonsillectomy: discriminant validation of the Postoperative Quality of Recovery Scale. *Acta anaesthesiologica Scandinavica* 2014; **58:** 345-51.
25. Liu J, Yuan W, Wang X, et al. Peripheral nerve blocks versus general anesthesia for total knee replacement in elderly patients on the postoperative quality of recovery. *Clin Interv Aging* 2014; **9:** 341-50.
26. Rasmussen LS, Steinmetz J Ambulatory anaesthesia and cognitive dysfunction. *Current opinion in anaesthesiology* 2015; **28:** 631-5.
27. Steinmetz J, Rasmussen LS Peri-operative cognitive dysfunction and protection. *Anaesthesia* 2016; **71 Suppl 1:** 58-63.
28. Brown Ct, Deiner S Perioperative cognitive protection. *British journal of anaesthesia* 2016; **117:** iii52-iii61.
29. Needham MJ, Webb CE, Bryden DC Postoperative cognitive dysfunction and dementia: what we need to know and do. *British journal of anaesthesia* 2017; **119:** i115-i25.

Table 1. Clinical outcomes of cardiac surgery patients with normal cognitive baseline scores

	Recovered at Day 1 *n = 79*	***Nonrecovered at Day 1*** *n = 167*	***Odds ratio*** *(95% CI)*	***P-value***
Length of stay				
Postoperative stay, days	8.0 ± 5.9	10.4 ± 10.0		0.02
<10 days, n (%)	69 (87)	124 (74)		
11 to 16 days, n (%)	8 (10)	24 (14)	2.4	0.03
>17 days, n (%)	2 (3)	19 (11)	(1.1-5.0)	
ICU stay, days	2.1 ± 3.0	3.2 ± 5.5		0.049
<2 days, n (%)	68 (86)	111 (66)		
3 to 5 days, n (%)	7 (9)	40 (24)	3.1	0.01
>6 days, n (%)	4 (5)	16 (10)	(1.5-6.3)	
Adverse outcomes				

Death, n (%)	3 (4)	10 (6)	1.6 (0.5-5.6)	0.47
Atrial fibrillation, n (%)	27 (34)	63 (38)	1.2 (0.7-2.0)	0.59
Myocardial infarction, n (%)	24 (30)	72 (43)	1.7 (1.0-3.1)	0.06
Stroke, n (%)	2 (3)	6 (4)	1.4 (0-3.1)	0.66
Acute kidney injury, n (%)	10 (13)	34 (20)	1.8 (0.8-3.7)	0.14
Normal function, n (%)	69 (87)	133 (80)		
Risk, n (%)	5 (6)	19 (11)		
Injury, n (%)	4 (5)	9 (5)	1.8	0.51

Failure, n (%)	1 (1)	3 (2)	(0.8-3.7)	
Loss of function, n (%)	0 (0)	0 (0)		
End-stage disease, n (%)	0 (0)	3 (2)		
Surgical site infection, n (%)	3 (4)	15 (9)	1.6 (0.7-3.6)	0.15
Other infection, n (%)	8 (10)	25 (15)	2.5 (0.7-8.3)	0.30

Status of recovery refers to the cognitive subdomain on the first postoperative day, and renal function is defined according to the RIFLE criteria. Odds ratios and confidence intervals are calculated as superiority of cognitive recovery to cognitive non-recovery at first postoperative day i.e. an odds ratio of more than 1 indicates superiority of recovery, whereas an odds ratio of less than 1 indicates superiority of non-recovery. Data reported as means ± standard deviations or absolute numbers and percentages of patients. ICU, intensive care unit; CI, confidence interval

Table 2. Clinical outcomes of patients undergoing cardiac surgery with low cognitive baseline scores

	Recovered at Day 1 *n = 89*	***Non-recovered at Day 1*** *n = 142*	***Odds ratio*** *(95% CI)*	***P-value***
Length of stay				
Postoperative stay, days	8.2 ± 4.7	12.0 ± 11.1		<0.01
<10 days, n (%)	75 (84)	89 (63)		
11 to 16 days, n (%)	9 (10)	30 (21)	3.2	<0.01
>17 days, n (%)	5 (6)	23 (16)	(1.7-6.2)	
ICU stay, days	2.4 ± 2.5	3.9 ± 4.7		<0.01
<2 days, n (%)	72 (81)	82 (58)		
3 to 5 days, n (%)	9 (10)	38 (27)	3.1	<0.01

>6 days, n (%)	8 (9)	22 (15)	(1.7-5.8)	
Adverse outcomes				
Death, n (%)	1 (1)	13 (9)	-	0.01
Atrial fibrillation, n (%)	33 (37)	51 (36)	1.0 (0.6-1.6)	0.86
Myocardial infarction, n (%)	33 (37)	58 (41)	1.2 (0.7-2.0)	0.57
Stroke, n (%)	1 (1)	2 (1)	-	0.85
Acute kidney injury, n (%)	19 (21)	30 (21)	0.99 (0.5-1.9)	0.97
Normal function, n (%)	70 (79)	112 (79)	0.99 (0.5-1.9)	0.98
Risk, n (%)	11 (12)	16 (11)		
Injury, n (%)	3 (3)	5 (4)		

Failure, n (%)	2 (2)	5 (4)		
Loss of function, n (%)	0 (0)	0 (0)		
End-stage disease, n (%)	3 (3)	4 (3)		
Surgical site infection, n (%)	1 (1)	12 (8)	-	0.02
Other infection, n (%)	7 (8)	32 (23)	3.4 (1.5-7.9)	<0.01

Status of recovery refers to the cognitive subdomain on the first postoperative day, and renal function is defined according to the RIFLE criteria. Odds ratios and confidence intervals are calculated as superiority of cognitive recovery to cognitive non-recovery at first postoperative day i.e. an odds ratio of more than 1 indicates superiority of recovery, whereas an odds ratio of less than 1 indicates superiority of non-recovery. Data reported as means ± standard deviations or absolute numbers and percentages of patients. ICU, intensive care unit; CI, confidence interval; -, odds ratio cannot be calculated due to low number of events

Figure Legends

Figure 1. The number of tests that were scored above the cut-off values for low baseline cognition are shown for the 11-item neuropsychological better (Y axis) and the PostopQRS (X axis). The cut off for low baseline cognition is ≤ 9 neuropsychological, or ≤4 PostopQRS tests.

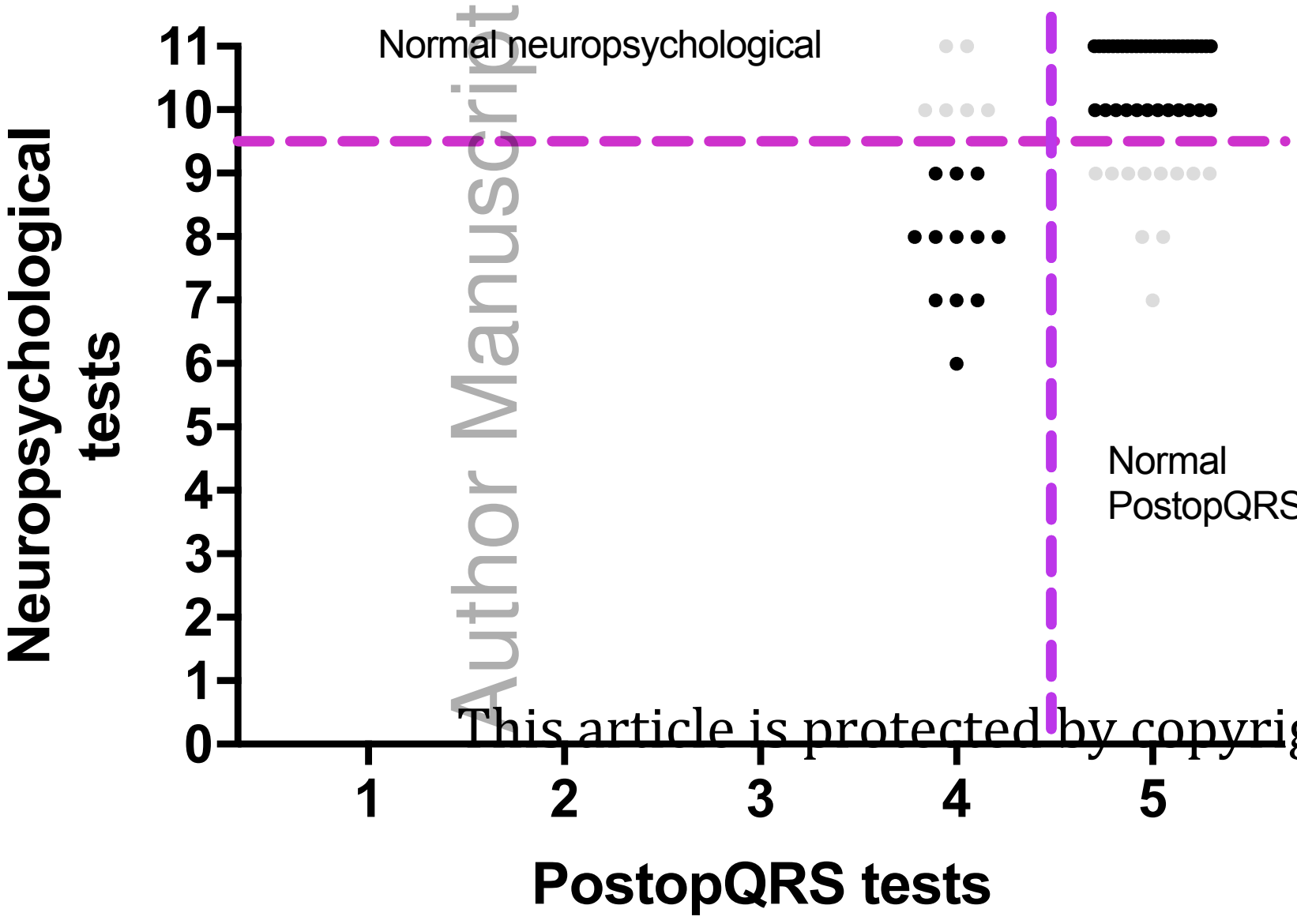
Figure 2. The proportion of participants who recovered in the PostopQRS cognitive domain is shown over time for participants with low and normal baseline cognition. Low baseline cognition is categorised by both PostopQRS and neuropsychological definitions.

Figure 3. The proportion of participants who recovered using neuropsychological criteria is shown over time for participants with low and normal baseline cognition. Low baseline cognition is categorised by both PostopQRS and neuropsychological definitions.

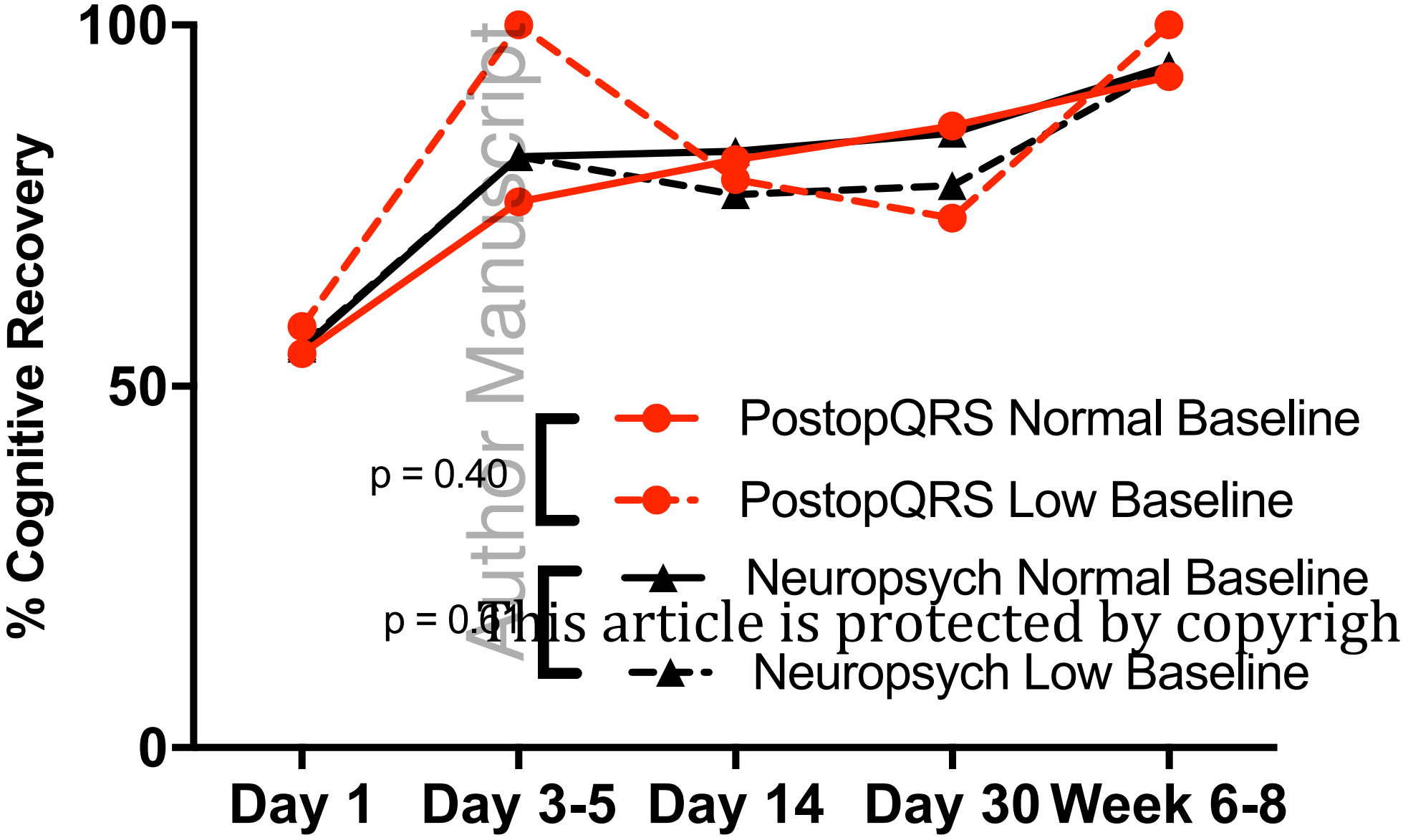
Figure 4. Recovery over time is shown for the SIRS sub study cohort in participants with normal baseline cognition, categorized by the presence or absence of cognitive recovery on Day 1 after surgery.

Figure 5. Recovery over time is shown for the SIRS sub study cohort in participants with low baseline cognition, categorized by the presence or absence of cognitive recovery on Day 1 after surgery.

Number of tests scoring normal at baseline



PostopQRS Cognitive Recovery



Neuropsychological Recovery