

Novel Approaches to Global Benchmarking of Risk-Adjusted Surgical Outcomes

Richard Trafford Spence (SPNRIC004) MBChB (Hons) MPhil PhD

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfillment of the requirements for the degree
Masters of Medicine (Surgery)

SUPERVISORS

Professor Eugenio Panieri MBChB FCS (SA)



The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

DECLARATION

I, *Richard T Spence*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:

Signed by candidate

Date: 31 July 2017

... the real need (in global health) is to close the data gaps, especially in low and middle-income countries, so that we no longer have to rely heavily on statistical modeling for data on disease burden.

Dr. Margaret Chan, Director General, World Health Organization ¹

Table of Contents

ACKNOWLEDGEMENTS AND CONTRIBUTIONS	5
LIST OF TABLES	6
LIST OF FIGURES	6
ABBREVIATIONS	7
ABSTRACT	8
MANUSCRIPT	10
<i>Derivation, Validation and Application of A Pragmatic Risk Prediction Index for Benchmarking of Surgical Outcomes</i>	10
INTRODUCTION	11
METHODS	12
RESULTS	15
DISCUSSION	17
REFERENCES	20
TABLES	22
FIGURES	27
APPENDIX	30
AMERICAN COLLEGE OF SURGEONS NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM ESSENTIALS WORKSHEET	30
HREC APPROVAL	33
DRC APPROVAL	34
REVIEWERS COMMENTS	35

Acknowledgements and contributions

This work is the result of a collaborative effort between the Codman Center for Clinical Effectiveness at Harvard University and the Department of Surgery at the University of Cape Town. My sincere thanks to Professor Matt Hutter at Massachusetts General Hospital for his support and to Professor David C Chang for his expertise and guidance. To my supervisor, Professor Eugenio Panieri, thank you for always looking for avenues to help advance my career and for your mentorship during this study. More specifically, the individuals involved in this manuscript and their contributions are outlined below;

Author contributions

Study design: RT Spence, DC Chang, MM Hutter, E Panieri

Data collection: RT Spence, GA Anderson, E Panieri

Manuscript writing: RT Spence, HMA Kaafarani, GA Anderson, E Panieri

Critical review and editing: RT Spence, MM Hutter, HMA Kaafarani, E Panieri

List of Tables

1. Table 1. Comparison of the preoperative risk characteristics in the ACS-NSQIP and SA datasets
2. Table 2. Occurrence of endpoints in the ACS-NSQIP and SA datasets
3. Table 3. The multivariate model used to derive the Codman Score presented in order of entry
4. Table 4 Components of the Codman Score

List of Figures

1. Figure 1. Stepwise process to derive, validate and apply the Codman Score
2. Figure 2. Comparison of the discriminatory ability to predict an in-hospital death in the ACS-NSQIP derivation and validation datasets
3. Figure 3. Mortality and morbidity rates by Codman Score

Abbreviations

American College of Surgeons' National Surgical Quality Improvement Program (ACS-NSQIP)
American Society of Anesthesia (ASA)
Area under the receiver-operator characteristic curve (ROC curve).
Current Procedure Code (CPT)
Goodness-of-fit statistic (GOF)
Groote Schuur Hospital (GSH)
Low income country (LIC)
Middle Income Country (MIC)
Observed versus expected (O/E) ratio.
Optimal cutoff point (C*)
Research Electronic Data Capture (REDCap)
South Africa (SA)

Abstract

Background

Despite the existence of multiple validated risk-assessment and quality benchmarking tools in surgery, their utility outside of High Income Countries is limited. We sought to derive, validate and apply a scoring system that is both 1) feasible, and 2) reliably predicts mortality in a Middle Income Country (MIC) context.

Methods

A 5-step methodology was used:

1. Development of a de novo surgical outcomes database modeled around the American College of Surgeons' National Surgical Quality Improvement Program (ACS-NSQIP) in South Africa (SA Dataset)
2. Use of the resultant data to identify all predictors of in-hospital death with more than 90% capture indicating feasibility of collection
3. Use these predictors to derive and validate an integer-based score that reliably predicts in-hospital death in the 2012 ACS-NSQIP
4. Apply the score in the original SA dataset and demonstrate its performance
5. Identify threshold cutoffs of the score to prompt action and drive quality improvement.

Results

Following Step one-three above, the 13 point Codman's score was derived and validated on 211,737 and 109,079 patients, respectively, and includes: 1) age \geq 65 (1), partially or completely dependent functional status (1), preoperative transfusions \geq 4 units (1), emergency operation (2), sepsis or septic shock (2) American Society of Anesthesia (ASA) score \geq 3 (3) and operative procedure (1-3).

Application of the score to 373 patients in the SA dataset showed good discrimination and calibration to

predict an in-hospital death. A Codman Score of 8 is an optimal cutoff point for defining expected and unexpected deaths.

Conclusion

We have designed a novel risk prediction score specific for a MIC context. The Codman Score can prove useful for both 1) preoperative decision-making and 2) benchmarking the quality of surgical care in MIC's.

Manuscript

Derivation, Validation and Application of A Pragmatic Risk Prediction Index for Benchmarking of Surgical Outcomes

Richard T Spence MBChB MPH^{1,2}, David C Chang MBA MPH PhD^{1,3}, Haytham MA Kaafarani MD MPH FACS^{1,3}, Eugenio Panieri MBChB FCS², Geoffrey A Anderson MD MPH³, Matthew M Hutter MD MPH FACS^{1,3}.

1. Codman Center for Clinical Effectiveness in Surgery, Department of General Surgery, Massachusetts General Hospital, Boston, USA.
2. Department of Surgery, University of Cape Town, Cape Town, South Africa.
3. Harvard Medical School, Boston, MA

Word Count: 2491

Article type: Original scientific article

Corresponding author:

Richard T Spence.

Email: spnric004@myuct.ac.za

Contact number: + 2721 8801112

Fax number: +2721 8801114

INTRODUCTION

A handful of landmark studies in the emerging field of surgical outcomes research have shown that there are significant variations in mortality following surgery between countries (1-3). National and international strategies to improve perioperative outcomes are urgently needed but reliable surgical outcomes data are lacking in the areas of the world with the highest surgical burden, greatest unmet need and likely the greatest variation in surgical outcomes (4).

To date in the United States, the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) remains the most robust risk-adjusted and reliable tool available and, most important, the only tool that is readily accepted by most surgeons (5). The ACS-NSQIP has been shown to reduce both morbidity and mortality in enrolled hospitals with initially worse performing hospitals having a greater likelihood of improvement (6,7). A significant problem with the ACS-NSQIP is that its expense limits the number of participating hospitals and excludes most smaller and rural hospitals in the United States- hospitals about which one might legitimately wish to ask certain quality and safety questions (5). Additionally, the program requires the retrospective collection of over 130 variables per patient further limiting the generalizability of the ACS-NSQIP. For these reasons, in it's current format, such a program has limited applicability in a MIC.

Despite the existence of multiple validated risk-assessment and quality benchmarking tools in surgery, their utility in low resource settings is often limited. We still lack a clinically meaningful objective metric that can be applied preoperatively to a general surgery cohort in a MIC context. The Surgical Apgar Score has been well validated globally but is based on data collected intra-operatively at 5-minute intervals, limiting its preoperative application (8).

Ernest Amory Codman was a courageous early 20th century champion for an “end results system” to track hospital outcomes and compare them between providers (9). Acknowledging the current global variation in patient outcomes following noncardiac surgery and the lack of meaningful objective metrics that reliably predict adverse events in more resource-limited settings, were the main drivers behind developing the Codman Score. Our primary aim was to develop a simple index with adequate precision to preoperatively identify patients that have a high risk of in-hospital death in a MIC context.

METHODS

There were five distinct steps in this study, as summarized in Figure 1.

Step 1: Development of a SA dataset

Groote Schuur Hospital (GSH) is a government-funded, tertiary teaching hospital situated in Cape Town, South Africa. It is the chief academic hospital of the University of Cape Town and the referral hospital for three secondary level surgical departments in the Western Cape Metropolitan. GSH has a bed capacity for 1000 patients and an estimated 10,000 operations are performed annually at GSH, the risk-adjusted outcomes of which are not formally audited or known. The patient population was defined by a systematic sample of major vascular and general surgery operations performed at GSH on adult patients that met inclusion criteria for the ACS-NSQIP Essentials Program. The details of this program have been well validated and published previously and the datasheet for this program has been included in the Appendix. Local adaptations are included in this text. The Departments of Surgery and Quality Assurance at GSH employed a clinical auditor for this study. A pilot period of one month was performed by the clinical auditor under direct supervision of the principle investigators where the main

focus was on data variable definitions and to identify and solve any logistical issues of case identification and data capture. After the 1-month pilot period, the clinical reviewer functioned independently for the 3-month recruitment period from 1st April to 30th June 2014.

Consecutive patients for inclusion were identified from the main theatre register that captures all operative activity in the hospital. Patients were then followed-up in the ward post-operatively where the data extraction occurred. A data collection sheet for this purpose was downloaded from the ACS website. After the one-month pilot, the datasheet was modified to increase local relevance for categories describing race, ethnicity, and preferred language. An online dataset with the final variables for inclusion was designed using REDCap (10). An iPad was given to the clinical reviewer so data extraction could occur directly at the bedside into the REDCap database. Once entered in the database, a 30-day follow-up from date of surgery by telephonic interview and chart review then took place. Primary outcome for the simple scoring system was in-hospital death. The secondary outcome was any morbidity at 30-days (including in-hospital death). This was defined by an occurrence of one of the specific postoperative complications spanning each of the organ systems defined in the ACS-NSQIP protocol.

Step 2: Identification of predictors

To ensure that the resultant scoring system would be feasible to apply, step two involved identifying only preoperative data variables collected in the SA dataset that were greater than 90% complete. The univariate association between these predictors and in-hospital death in the SA dataset was tested with significance set at $p=0.1$ to err on the side of model inclusion. Step two therefore involved identifying variables in the SA dataset that were both 1) feasible to collect and 2) potentially associated with in-hospital death.

Step3: Derivation and validation of a simple scoring index

The 2012 sample of 320,816 patients from the ACS-NSQIP essentials program was used to derive and validate the simple scoring system. This dataset was chosen, and the methodology of data acquisition was replicated in South Africa, due to the breadth of preoperative variables collected for each patient to enhance the exploratory nature of this study and increase the generalizability of the derived scoring system. Two thirds of the cohort were randomly assigned to the derivation cohort (n=211,737) and one third was to the validation cohort (n=109,079). Operative categories were defined by CPT code according to a classification system previously described in the Patient Safety in Surgery Study (11). Potentially significant predictors identified in the SA dataset (in step 2) were then dichotomised to further simplify the resultant score at thresholds cited in the literature. These binary predictors were entered into a logistic regression analysis in the derivation dataset using a manual stepwise forward selection algorithm by a process described by Anderson et al (12). An a priori decision was made to include procedure category as the first variable in the model. The first step was to perform simple logistic regression and identify the binary predictor to add to the procedure category model with the highest discriminatory ability to predict in-hospital death as assessed by the area under the receiver-operator characteristic curve (ROC curve). The ROC curve is used to define the c-statistic, a discriminative measure to identify how well a model separates two groups (i.e. in this case, patients who did and did not die in hospital) (13). For each additional step, remaining variables were individually added to the algorithm and the resultant model combination with the highest c-statistic would progress to the next step until further variable addition resulted in minimal gain in discriminatory ability. The beta coefficients of the final binary predictors and the individual operative categories were then divided and rounded to the nearest integer to create a scoring system (14). Logistic regression was then repeated using this single score to predict in-hospital death and any morbidity at 30 days in the validation dataset.

The ROC curves of the score derived in the derivation and validation datasets were compared. A probability of in-hospital death and any morbidity was then generated for each value of the validated Codman Score.

Step 4: Application of the Codman Score

The validated scoring rubric was then applied to the SA dataset. The performance of the Codman Score was assessed by both its discriminatory ability as well as its ability to calibrate in-hospital deaths and any morbidity. Measure of calibration of the Codman Score was assessed by the Hosmer and Lemeshow goodness-of-fit statistic (GOF) (13).

Step 5: Identify threshold cutoffs of the score to prompt action and drive quality improvement

A threshold score was identified in the ACS-NSQIP database with the highest sensitivity and specificity to identify an in-hospital death. This was identified using the Youden index (ie. J-statistic), a statistical tool that defines the optimal cutoff point (C*) for the differentiating ability of a test using sensitivity and specificity of that test (15). This C* was applied to the SA dataset to objectively classify patients into four categories; expected successes, unexpected failures, unexpected successes and expected failures.

Ethics was granted by the University of Cape Town's research ethics committee. All analyses were performed using STATA 14.

RESULTS

A total of 373 patients were included in the SA dataset. The mean age of the cohort was 49.9 years (range 13 – 91), 57.3% were female, 20.3% were vascular patients and 79.6% were general surgery patients. The patients in the two datasets were not similar as demonstrated by the comparison in Table 1 based on the preoperative predictors that were greater than 90% complete in the SA dataset. There was a greater proportion of patients classed as an ASA greater or equal to 3 in the ACS-NSQIP

dataset. Patients in the SA dataset had a higher proportion of emergency cases, diabetics, symptomatic dyspnea, functional dependency, congestive heart failure, chronic obstructive pulmonary disease, acute renal failure, disseminated cancer and preoperative sepsis. Higher rates of vascular and endocrine surgery was performed in the SA dataset, whereas, higher rates of hernia, thoracic and lymphatic surgery were performed in the ACS-NSQIP.

The occurrence of endpoints, including outcomes for prediction, in both the ACS-NSQIP and SA datasets are presented in Table 2. The SA dataset had a significantly higher proportion of all outcome measures except for CNS occurrences. The unadjusted in-hospital mortality rate was 6.4% (95% CI 4.34 – 9.44) and 30-day any complication rate was 34.1% (95% CI 29.73 – 40.26) in the SA dataset. In the ACS-NSQIP dataset the unadjusted in-hospital mortality rate was 1.1% (95%CI 1.04 – 1.11) and 30-day any complication rate was 13.8% (95% CI 13.68 – 13.92).

The final multivariate model from the derivation dataset is presented in Table 3. The binary predictors are presented in order of entry into the model along with the c-statistics derived after addition. The resultant Codman Score relies on seven simple preoperative variables; the American Society of Anesthesia (ASA) score, emergency status, the amount of packed red blood cells transfused within 72 hours prior to surgery, age, sepsis status, functional dependency status and surgical procedure. The total Codman Score ranges from 0 to 13, with higher scores indicating patients at higher risk of in-hospital death. The scoring rubric for the Codman Score is presented in Table 4.

There was no difference in the discriminatory ability of the score in the derivation and validation datasets ($p=0.16$) as seen by the ROC curve comparison in figure 2. When applied to the whole ACS-

NSQIP, there was a stepwise increase in both in-hospital mortality and any morbidity at 30 days as seen in figure 3. When applied to the SA dataset, 351 patients (94.1%) had all the data variables complete to compute a Codman Score of which the median score was 4 (IQR 3-6). The score had a c-statistic of 0.8672 (95% CI 0.8136 – 0.9209) and GOF statistic of 1.11 (p = 0.292) to predict an in-hospital death and a c-statistic of 0.7183 (95% CI 0.6496 – 0.7870) and GOF statistic of 0.04 (p = 0.848) to predict any complication at 30 days. In the ACS-NSQIP a C* of 8 was identified as the optimum threshold with the highest sensitivity (88.76%) and specificity (80.63%) to preoperatively identify an in-hospital death. Applied to the SA dataset, the C* identified 8 unexpected deaths, 30 unexpected survivors, 14 expected deaths and 300 expected survivors.

DISCUSSION

This study describes the derivation, validation and demonstrated utility of a simple 7-variable index, the Codman Score, designed to preoperatively identify adult general or vascular surgery patients at risk for in-hospital death in a MIC context.

Developed to ensure feasibility of data capture, the score can be used for preoperative planning as well as an objective surgical quality metric. This algorithm has been derived and validated on the ACS-NSQIP essentials dataset. Therefore, using the expected probabilities of mortality and morbidity for each score (0-13), surgical units around the world can use the simple and parsimonious Codman Score for global benchmarking of risk-adjusted outcomes against the ACS-NSQIP consortium by an objective “observed versus expected” quality metric (O/E).

Previous work has supported the fact that a model based on only a few variables may provide enough discrimination to measure surgical outcomes in a risk-adjusted manner. Rubinfeld et al found the c-statistic for mortality decreased only slightly from 0.907 using all variables to 0.902 using 10 variables

and argue that only a few variables are required for predictive accuracy (16). Dimick et al found that limited models based on 5 or 12 variables had comparable discrimination to a 21-variable model (17). Birkmeyer et al also found high correlation between a 5-variable and a 20-variable morbidity risk model and recommended that the new version of the NSQIP have no more than 5 to 10 core covariates (12). Anderson et al provide several examples of risk-adjustment models that may be appropriate for hospitals in resource-limited settings. Their 6-variable model with the highest discriminatory ability included ASA class, age, sepsis, functional status, cancer and emergency status (12). Our work has reproduced similar findings but has further consolidated the model into a pragmatic index, which also includes a procedure variable, simultaneously provides a preoperative risk prediction tool and has been validated outside the ACS NSQIP.

Clinical outcomes following major surgery are poorly described at a national level and even an institutional level in MIC's (4). Furthermore, by only reporting unadjusted postoperative mortality rate (POMR), as suggested by the Lancet Commission on Global Surgery, hospitals that manage sicker patients would appear to have worse outcomes (18). A more meaningful quality metric is needed. Such a metric for measuring quality is the observed versus expected (O/E) ratio. The O/E ratio is a measure of the degree of agreement between the predicted outcome (E) and the actual outcome (O) (18). With these data, meaningful comparison of risk-adjusted outcomes between surgical units can be achieved. Nevertheless, the adoption of the O/E ratio as a metric for surgical quality has been limited, mostly because the measurement of the 'E' in the O/E ratio has remained a challenge for surgical units outside quality consortia like the ACS-NSQIP. However, based on seven easily attainable preoperative variables, the Codman Score generates an expected individual probability for both mortality and morbidity.

The major limitation of this study is the small, single-centered SA dataset. Applying the ACS-NSQIP essentials methodology for a longer time was not feasible and would have threatened the accuracy of the dataset. Data quality was chosen over quantity for this study. One of the most accepted predictors of a poor outcome is a high ASA class, which is subject to inter-rater reliability (20). A study by Cohen ME et al. explored the effect of subjective preoperative variables on risk-adjusted assessment of hospital morbidity and mortality (21). Using ASA and functional health status, the authors showed that these variables have unique contributions risk-adjustment models and have little indication that they are subject to an important level of institutional bias. They therefore concluded that it is appropriate that they be used to assess risk-adjusted surgical quality (21). The inclusion of the blood transfusion requirement of 4 units or greater in the score may exclude the applicability of the Codman Score for resource limited settings where blood transfusion is not an option. Our model building process ranked this variable as the third most discriminatory after ASA and emergency status. The work by Anderson et al proposes the use of haematocrit as an alternative but this too may not be a practical to collect (12). The Lancet Commission included a safe and affordable blood supply as one of the ten essential needs for the provision of safe surgical and anaesthesia and if a surgical collaborative is interested in auditing their outcomes, access to a safe blood supply must be a priority step towards quality improvement. This is a model, which may be feasible in a low income country (LIC) or may require further modifications. It is necessary however, that there is external validation of the model in a LIC. Furthermore, to extend the model beyond general and vascular surgery, it will need possible inclusion of other surgical procedures into the model.

We believe this study provides evidence supporting the use of the Codman Score and justifies a prospective interrogation of its accuracy and generalizability. The Codman Score is easily calculated, predictive and discriminative for major complications among adults undergoing in-patient, general surgery operative procedures. Its application has the potential to advance surgical quality improvement initiatives in MIC's.

REFERENCES

- (1) Haider AH, Hashmi ZG, Gupta S, Zafar SN, David JS, Efron DT, et al. Benchmarking of trauma care worldwide: the potential value of an International Trauma Data Bank (ITDB). *World J Surg* 2014 Aug;38(8):1882-1891.
- (2) Pearse RM, Moreno RP, Bauer P, Pelosi P, Metnitz P, Spies C, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012 Sep 22;380(9847):1059-1065.
- (3) Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009 Jan 29;360(5):491-499.
- (4) Meara JG, Greenberg SL. Global surgery as an equal partner in health: no longer the neglected stepchild. *Lancet Glob Health* 2015 Apr 27;3 Suppl 2:S1-2.
- (5) Farmer DL. NSQIP lite: a potential tool for global comparative effectiveness evaluations. *Arch Surg* 2012 Sep;147(9):803-804.
- (6) Hall BL, Hamilton BH, Richards K, Bilimoria KY, Cohen ME, Ko CY. Does surgical quality improve in the American College of Surgeons National Surgical Quality Improvement Program: an evaluation of all participating hospitals. *Ann Surg* 2009 Sep;250(3):363-376.
- (7) Cohen ME, Liu Y, Ko CY, Hall BL. Improved Surgical Outcomes for ACS NSQIP Hospitals Over Time: Evaluation of Hospital Cohorts With up to 8 Years of Participation. *Ann Surg* 2015 Feb 26.
- (8) Haynes AB, Regenbogen SE, Weiser TG, Lipsitz SR, Dziekan G, Berry WR, et al. Surgical outcome measurement for a global patient population: validation of the Surgical Apgar Score in 8 countries. *Surgery* 2011 Apr;149(4):519-524.
- (9) Berwick DM. Measuring surgical outcomes for improvement: was Codman wrong? *JAMA* 2015 Feb 3;313(5):469-470.

- (10) Project REDCap. 2015; Available at: <http://www.project-redcap.org/>. Accessed June, 24, 2015.
- (11) Johnson RG, Arozullah AM, Neumayer L, Henderson WG, Hosokawa P, Khuri SF. Multivariable predictors of postoperative respiratory failure after general and vascular surgery: results from the patient safety in surgery study. *J Am Coll Surg* 2007 Jun;204(6):1188-1198.
- (12) Andersen J, Lassiter R, Bickler S, Talamini M, Chang D. Brief tool to measure risk-adjusted surgical outcomes in resource-limited hospitals. *Arch Surg* 2012;147(9):798-803.
- (13) Cook F. Analytical epidemiology lecture series. Harvard TH Chan School of Public Health, Program in Clinical Effectiveness: <https://www.hsph.harvard.edu/clinical-effectiveness/2014>.
- (14) Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* 1999 Sep 7;100(10):1043-1049.
- (15) Batterton KA, Schubert CM. A nonparametric fiducial interval for the Youden index in multi-state diagnostic settings. *Stat Med* 2015 Aug 16.
- (16) Rubinfeld I, Farooq M, Velanovich V, Syed Z. Predicting surgical risk: how much data is enough? *AMIA Annu Symp Proc.* 2010;2010:777-781.
- (17) Dimick JB, Osborne NH, Hall BL, Ko CY, Birkmeyer JD. Risk adjustment for comparing hospital quality with surgery: how many variables are needed? *J Am Coll Surg.*2010;210(4):503-508.
- (18) Ariyaratnam R, Palmqvist CL, Hider P, Laing GL, Stupart D, Wilson L, et al. Toward a standard approach to measurement and reporting of perioperative mortality rate as a global indicator for surgery. *Surgery* 2015 Jul;158(1):17-26.
- (19) Henderson WG, Daley J. Design and statistical methodology of the National Surgical Quality Improvement Program: why is it what it is? *Am J Surg* 2009 Nov;198(5 Suppl):S19-27.
- (20) Haynes SR, Lawler PG. An assessment of the consistency of ASA physical status classification allocation. *Anaesthesia* 1995 Mar;50(3):195-199.
- (21) Cohen ME, Bilimoria KY, Ko CY, Richards K, Hall BL. Effect of subjective preoperative variables on risk-adjusted assessment of hospital morbidity and mortality. *Ann Surg* 2009 Apr;249(4):682-689.

TABLES

Table 1. Comparison of the preoperative risk characteristics in the ACS-NSQIP and SA datasets

Preoperative risk characteristic	ACS-NSQIP (n=320,815) N (%)	SA Dataset (n=373) N (%)	Odds ratio for in-hospital death and 95% CI's in the SA dataset
Age category ≥ 65	143,212 (44.64)	80 (21.45)	0.932 (0.788 - 1.103)
Male gender	140,917 (43.92)	159 (42.63)	2.231 (1.299 - 3.833)
Non-white race	79,565 (24.80)	331 (88.74)	0.941 (0.682 - 1.298)
Emergency case	45,585 (14.21)	137 (36.73)	5.333 (2.994 - 9.501)
ASA $\geq 3^*$	152,739 (47.78)	103 (33.76)	2.551 (1.468 - 4.433)
Speciality vascular	42,900 (13.37)	76 (20.38)	0.971 (0.502 - 1.877)
Diabetic	52,676 (16.42)	79 (21.18)	1.017 (0.526 - 1.953)
Smoking within the year	61,373 (19.13)	193 (51.74)	1.043 (0.588 - 1.849)
Dyspnoea \geq moderate exertion	23,098 (7.20)	126 (36.63)	2.689 (1.523 - 4.749)
Functional status \geq partially dependent	9,700 (3.02)	55 (14.75)	2.705 (1.401 - 5.227)
Ventilator dependent w/in 48hrs	1,673 (0.52)	2 (0.54)	4.016 (0.248 - 65.108)
Chronic Obstructive Pulmonary Disease	15,845 (4.94)	24 (6.43)	3.306 (1.326 - 8.239)
Congestive Heart Failure	2,640 (0.82)	8 (2.14)	7.069 (1.642 - 30.429)
Hypertension	148,458 (46.28)	134 (38.29)	0.802 (0.448 - 1.436)
Acute Renal Failure	1,663 (0.52)	16 (4.29)	4.872 (1.697 - 13.987)
Dialysis	6,525 (2.03)	6 (1.61)	6.098 (0.997 - 37.301)
Cancer	7,364 (2.30)	23 (6.17)	2.652 (1.048 - 6.709)
Open wound	13,671 (4.26)	37 (9.92)	2.001 (0.922 - 4.341)
Long term steroid use	12,275 (3.83)	12 (3.22)	1.352 (0.045 - 2.778)
>10% weight loss	5,837 (1.82)	25 (6.70)	1.721 (0.681 - 4.349)
Preoperative blood transfusion ≥ 4 packed cells in 72 hours	4,456 (1.39)	9 (2.41)	8.702 (2.113 - 35.838)
Septic status \geq sepsis	11,524 (3.59)	55 (15.67)	4.811 (2.505 - 9.239)
Procedure categories (CPT range)			
1. Integumentary and Musculoskeletal: Integumentary and musculoskeletal system (10000-29999)	40, 538 (12.64)	44 (11.8)	
2. Vascular A: Thoracic and thoracoabdominal aneurysms, embolectomy/thrombectomy, venous reconstruction, and endovascular repair (33001-34900)	2, 777 (0.87)	31 (8.31)	
3. Vascular B: Aneurysm, blood vessel repair, thromboendarterectomy, angiосcopy, angioplasty and atherectomy, bypass and	6, 591 (2.05)	30 (8.04)	

composite grafts, other artery, and vein (35001-37799)			2.022 (1.799 – 2.336)
4. Thoracic and lymphatic: Respiratory system, hemic and lymphatic systems, mediastinum, and diaphragm (30000-32999, 38000-39999)	35, 501 (11.07)	3 (0.8)	
5. Upper Gastrointestinal: Mouth, palate, salivary glands, pharynx, adenoids, and esophagus (40000-43499)	5, 317 (1.66)	5 (1.34)	
6. Mid and lower Gastrointestinal: Stomach, intestines, appendix and mesentery, rectum and anus, liver, biliary tract, pancreas, abdomen, peritoneum, and omentum (nonhernia) (43500-49429, 49650-49999)	172, 539 (53.78)	181 (48.53)	
7. Hernia surgery: Hernioplasty, herniorrhaphy, herniotomy (49491-49611)	43, 682 (13.62)	25 (6.7)	
8. Endocrine: Endocrine system (60000-60999)	13, 871 (4.32)	54 (14.48)	

Table 2. Occurrence of endpoints in the ACS-NSQIP and SA datasets

Occurrences	ACS-NSQIP (N=320,816)		SA dataset (N=373)		P-value
	N	%	N	%	
Wound	15,520	4.84	27	7.24	<0.05
Respiratory	10,061	3.14	19	5.09	<0.05
Urinary tract	6,643	2.07	21	5.63	<0.05
CNS	749	0.23	1	0.27	NS
Cardiac	2,492	0.23	9	2.41	<0.05
Other	27,371	8.53	50	13.40	<0.05
Outcomes for prediction					
Any morbidity at 30 days	44,289	13.81	127	34.05	<0.05
In-hospital death	3,454	1.08	24	6.43	<0.05

Table 3. The multivariate model used to derive the Codman Score
presented in order of entry

Procedure category	Beta-Coefficient	P-Value	ROC of model after variable addition
Endocrine	Reference		
Hernia surgery	1.184	<0.05	0.5642
Integumentary and Musculoskeletal	1.738	<0.01	
Thoracic and Lymphatic	2.403	<0.0001	
Upper Gastrointestinal	2.125	<0.001	
Mid and Lower Gastrointestinal	2.424	<0.0001	
Vascular A	2.934	<0.0001	
Vascular B	2.769	<0.0001	
Binary predictor			
ASA ≥ 3	2.849	<0.0001	0.8053
Emergency case	1.513	<0.0001	0.8751
Preoperative blood transfusion ≥ 4 packed cells in 72 hours	1.105	<0.0001	0.8813
Age category ≥ 65 years	0.958	<0.0001	0.9009
Septic status \geq sepsis	1.562	<0.0001	0.9079
Functional status \geq partially dependent	0.921	<0.0001	0.9152
Model with all 23 variables			0.9116

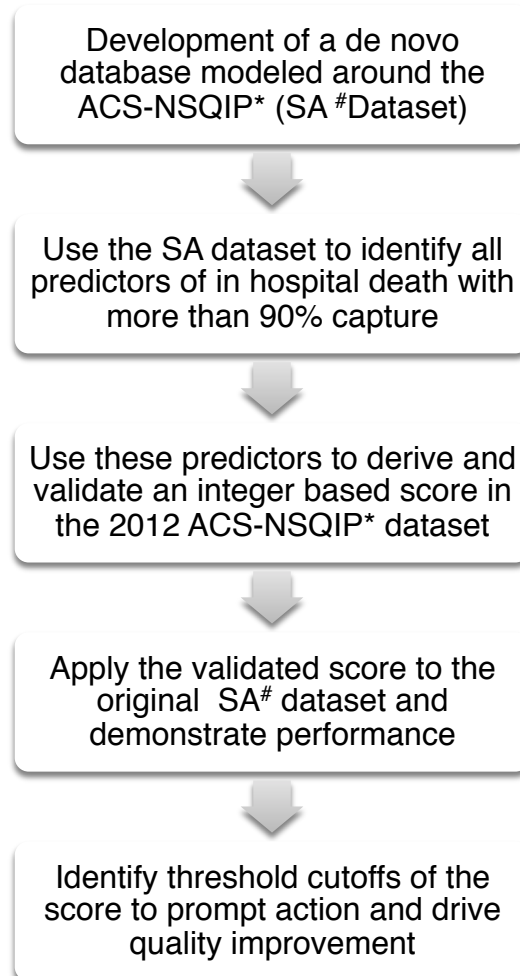
Table 4 Components of the Codman Score

Preoperative component	Score
Age (years)	
12-65	0
65 or older	1
Functional dependency status	
Independent	0
Partially or fully dependent	1
Preoperative blood transfusion within 72 hours of surgery	0
Less than 4 packed red blood cells	1
4 or more packed red blood cells	
Emergency status	0
Elective operation	2
Emergency operation	
Preoperative sepsis status	
None or SIRS*	0
Sepsis or septic shock	2
ASA Score	
Less than 3	0
3 or greater	3
Preoperative score	0 – 10
Operative procedure	
Endocrine/ Hernia	1
Gastro-intestinal / Thoracic/ Lymphatic/ Integumentary/ Musculoskeletal	2
Vascular	3
Procedure score	1-3
Total score	1-13

SIRS* Systemic Inflammatory Response Syndrome

FIGURES

Figure 1. Stepwise process to derive, validate and apply the Codman Score



ACS-NSQIP American College of Surgeons National Surgical Quality Improvement Program
SA # South African*

Figure 2. Comparison of the discriminatory ability to predict an in-hospital death in the ACS-NSQIP
derivation and validation datasets

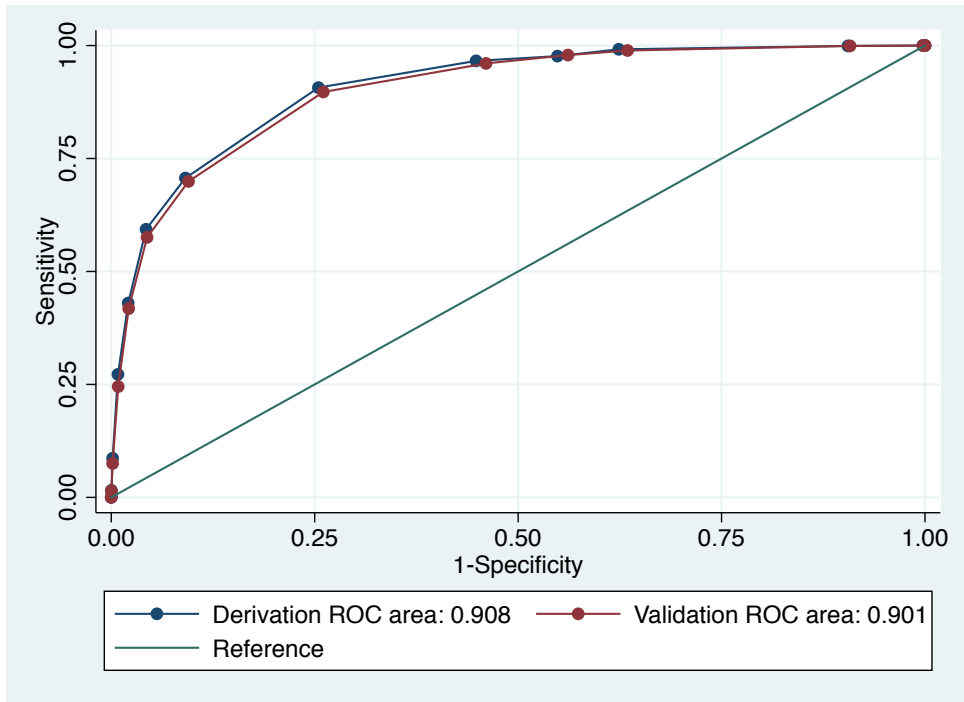
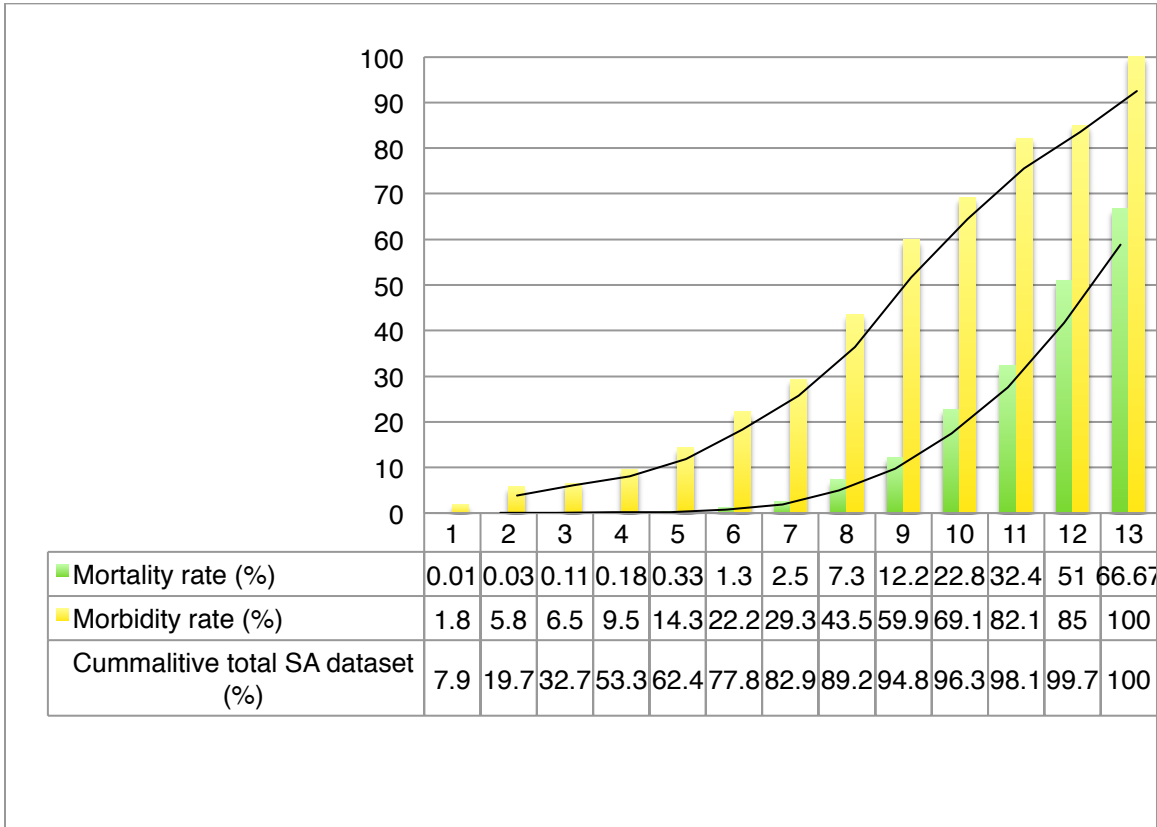


Figure 3. Mortality and morbidity rates by Codman Score



Appendix

AMERICAN COLLEGE OF SURGEONS NATIONAL SURGICAL QUALITY IMPROVEMENT PROGRAM ESSENTIALS WORKSHEET

*IDN _____ Cycle Number _____
LMRN _____ Case Number _____

Last Name: _____ First: _____ MI: _____
Street Address: _____
City/Town: _____ State/Province: _____ Zip: _____ Country: _____
Home Phone (____) _____ Work Phone (____) _____ Cell Phone (____) _____
*DOB: ____/____/____ (mm/dd/yyyy) Gender: Male Female
Race: White American Indian / Alaska Native Asian
 Black / African American Native Hawaiian / Other Pacific Islander Unknown
Ethnicity: Hispanic - YES NO Preferred Language: ENGLISH SPANISH

Principal Procedure _____ CPT Code _____

Patient Status: Inpatient Outpatient Elective Surgery: YES NO Unknown

Transfer / Origin Status:

Not transferred, admitted directly from home Transfer from other (i.e. Spinal Cord Injury Unit or other facility not listed)

Acute Care Hospital (inpatient status only) Transfer from outside Emergency Department

Nursing Home/Chronic Care Facility/Intermediate Care Unit

Unknown (if transferred from unknown location or Facility)

Hospital Admission Date: ____/____/____ *Operation Date: ____/____/____

Anesthesia Technique:

General

Spinal

Epidural

Regional

Local

MAC

None

Other

Unknown

*Surgical Specialty: (select one)

1. General Surgery 3. Thoracic 5. Orthopedics 7. Urology 9. Plastics

2. Vascular 4. Cardiac 6. Neurosurgery 8. Otolaryngology (ENT) 10. Gynecology

Attending Surgeon's Name: _____ Attending Surgeon's IDN: _____

LCN: _____ Encounter Number: _____

Revision: January 1, 2011 - 2 - ACS NSQIP ESSENTIALS

GENERAL RENAL

Height _____ Inches CM Acute Renal Failure w/in 24 hrs YES NO

Weight _____ Pounds KG Currently requiring or on Dialysis w/in 2 wks YES NO

Diabetes Mellitus Non-

Insulin Insulin NO NUTRITIONAL/IMMUNE/OTHER

Current Smoker w/in 1 year YES NO Disseminated Cancer YES NO

Dyspnea Mod.

Exertion At Rest NONE Open Wound (w/ or w/out infection) YES NO

Functional Health Status prior to

surgery I ___ PD ___ TD ___ Unk ___ Steroid use for chronic condition YES NO

PULMONARY >10% loss of body wt. last 6 months YES NO

Vent. Dependent w/in 48 hrs YES NO Bleeding disorders YES NO

COPD (severe) YES NO Preop Transfusions (RBC units w/in 72 hrs) YES NO

HEPATOBIILIARY Sepsis w/in 48 hours SIRS NO

Ascites w/in 30 days YES NO Sepsis

CARDIAC Sep Shock

CHF w/in 30 days YES NO

Hypertension req. meds. YES NO

LABORATORY DATA: (report preop lab values closest to the Procedure/Surgery start date & time)

Preop values should be within 90 days prior to surgery

PREOPERATIVE LAB DATA Value 90 days unknown Date

Serum Sodium (Na)

____/____/____
 Blood Urea Nitrogen (BUN) â
 ____/____/____
 Creatinine (Cr) â
 ____/____/____
 Albumin (ALB) â
 ____/____/____
 Total Bilirubin (TB) â
 ____/____/____
 Serum Glutamic-Oxaloacetic Transaminase (SGOT)/(AST) â
 ____/____/____
 Alkaline Phosphatase (Alk Phos) â
 ____/____/____
 White Blood Count (WBC) â
 ____/____/____
 Hematocrit (Hct) â
 ____/____/____
 Platelets (Plt) â
 ____/____/____
 Prothrombin Time (PT) â
 ____/____/____
 International Normalized Ratio (INR) â
 ____/____/____

Emergency Case: â YES â NO
 Wound Classification: â Clean â Clean/Contaminated â Contaminated â Dirty/Infected
 ASA Class (circle one): 1 2 3 4 5 6 None Assigned (for local anes. only)
 OPERATIVE TIMES: Procedure / Surgery Start: _____:_____ Procedure/Surgery Finish: _____:_____
 Revision: January 1, 2011 - 3 - ACS NSQIP ESSENTIALS

POSTOPERATIVE OCCURRENCES: â YES â NO
 (Although not required for this program, you may wish to document 'treatment' and 'outcome to date' of the occurrence for internal quality monitoring)

Date Treatments / Outcomes / Comments
 Wound Occurrences
 Superficial Incisional SSI ____/____/____
 Present at Time of Surgery? â YES â NO
 Deep Incisional SSI ____/____/____
 Present at Time of Surgery? â YES â NO
 Organ/Space SSI ____/____/____
 Present at Time of Surgery? â YES â NO
 Wound Disruption ____/____/____
 Respiratory Occurrences
 Pneumonia (PNA) ____/____/____
 Present at Time of Surgery? â YES â NO
 Unplanned Intubation ____/____/____
 Pulmonary Embolism ____/____/____
 On ventilator > 48 hours ____/____/____
 Present at Time of Surgery? â YES â NO
 Urinary Tract Occurrences
 Progressive Renal Insufficiency ____/____/____
 Acute Renal Failure ____/____/____
 Urinary Tract Infection (UTI) ____/____/____
 Present at Time of Surgery? â YES â NO
 CNS Occurrences
 Stroke / CVA ____/____/____
 Cardiac Occurrences
 Cardiac Arrest req. CPR ____/____/____
 Myocardial Infarction ____/____/____
 Other Occurrences
 Bleeding Requiring Transfusion (72h of surgery start time)
 (transfusion of 1-200 units) ____/____/____ # of units transfused: _____
 Deep Vein Thrombosis (DVT) req. Therapy ____/____/____
 Sepsis: Sepsis ____/____/____
 Septic Shock ____/____/____
 Other Postoperative Occurrences (ICD-9 code): ____/____/____ (ICD-9 code) _____

Discharge Destination:

â Chronic Care Facility, not Home â Home â Expired
â Unskilled Facility, not Home â Separate Acute care (transferred to another acute care facility) â Unknown
â Facility which was Home â Rehab

Post-op ICD.9 Code _____ Diagnosis: _____
Other Procedure CPT

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Concurrent Procedure CPT

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.
- 7.

Revision: January 1, 2011 - 4 - ACS NSQIP ESSENTIALS

Readmission:

Readmission for any reason within 30 days of the principle procedure? â YES â NO If yes, date: ____/____/____

Information Source (select one) â Medical Record â Patient/Family Report â Other

Was this readmission unplanned at the time of the principle procedure? â YES â NO

Select the primary reason for the unplanned readmission from the postoperative occurrences:

Superficial SSI Pulmonary Embolism Coma > 24 hours Deep Vein Thrombosis (DVT)

Deep SSI On ventilator > 48 hours Peripheral Nerve Injury Sepsis

Organ / Space SSI Progressive Renal Insufficiency Cardiac Arrest req CPR

Wound Disruption Acute Renal Failure Myocardial Infarction

Pneumonia Urinary Tract Infection (UTI) Bleeding Requiring Transfusion Other: ICD-9 code _____

Unplanned Intubation Stroke / CVA Graft / Prosthesis / Flap Failure

Notes ~ If ICD-9 is unknown, describe the reason.

Was this readmission for a post operative occurrence likely related to the principle surgical procedure? â YES â NO

Still in hospital > 30 days: â YES â NO

Hospital Discharge Date: ____/____/____ (mm/dd/yyyy)

Postoperative Death:

Postop Death w/in 30 days: â YES â NO Postop Death > 30 days:

(if remained in acute care)

â YES â NO

Date of death: ____/____/____ â Unknown Date of death: ____/____/____ â Unknown

Unplanned Reoperation:

Unplanned return to the operating room for a surgical procedure w/in 30 day postoperative period? â YES â NO

Was the return to the OR for a postop occurrence likely related to the principle procedure, or to any additional surgery performed under the same anesthetic as the principle procedure? â YES â NO

If yes, Surgery Date ____/____/____ CPT code _____ ICD9 code _____

Source (select one) â Medical Record â Patient/Family Report â Other

Notes ~ If CPT code is not documented, describe the surgery.

Was there a SECOND unplanned reoperation within 30 days? â YES â NO

Was the second return to the OR for a postoperative occurrence that was likely related to the principle procedure, or to any additional surgery (i.e., 'other' or 'concurrent') performed under the same anesthetic as the principle procedure? â YES â NO

If yes: Surgery Date ____/____/____ CPT code _____ ICD9 code _____

Source (select one) â Medical Record â Patient/Family Report â Other

Notes ~ If CPT code is not documented, describe the surgery.

Were there more than two unplanned reoperations for an adverse outcome related to the principal surgery within 30 days?

â YES â NO

HREC APPROVAL



UNIVERSITY OF CAPE TOWN
Faculty of Health Sciences
Human Research Ethics Committee



Room E53-46 Old Main Building
Groote Schuur Hospital
Observatory 7925
Telephone [021] 406 6626
Email: shurette.thomas@uct.ac.za
Website: www.health.uct.ac.za/fhs/research/humanethics/forms

26 May 2017

HREC REF: 335/2017

Prof E Panieri
Surgery
J-floor, OMB

Dear Prof Panieri

PROJECT TITLE: NOVEL APPROACHES TO GLOBAL BENCHMARKING OF RISK-ADJUSTED SURGICAL OUTCOMES BEYOND POST OPERATIVE MORTALITY RATE (MMED CANDIDATE - DR R SPENCE) SUB-STUDY LINKED TO 338/2014

Thank you for submitting your study to the Faculty of Health Sciences Human Research Ethics Committee.

It is a pleasure to inform you that the HREC has **formally approved** the above-mentioned study subject to the annual approval of the study 338/2014.

Approval is granted for one year until the 30 May 2018.

Please submit a progress form, using the standardised Annual Report Form if the study continues beyond the approval period. Please submit a Standard Closure form if the study is completed within the approval period.

(Forms can be found on our website: www.health.uct.ac.za/fhs/research/humanethics/forms)

Please quote the HREC REF in all your correspondence.

Please note that the ongoing ethical conduct of the study remains the responsibility of the principal investigator.

Please note that for all studies approved by the HREC, the principal investigator **must** obtain appropriate Institutional approval before the research may occur.

The HREC acknowledge that the student, Dr Richard Spence will also be involved in this study.

Yours sincerely

PROFESSOR M BLOCKMAN
CHAIRPERSON, FHS HUMAN RESEARCH ETHICS COMMITTEE

Federal Wide Assurance Number: FWA00001637.
Institutional Review Board (IRB) number: IRB00001938

HREC 335/2017

DRC APPROVAL



UNIVERSITY OF CAPE TOWN



Department of Surgery
Departmental Research Committee
Dr Timothy Pennel
D24 Office, Grootte Schuur Hospital
Observatory 7925
South Africa
Tel (021) 404 3430
Email: tim.pennel@uct.ac.za

19 May 2017

Dr R Spence
Department of Surgery
University of Cape Town

Dear Dr Spence
RE: Project 2017/061

PROJECT TITLE: Novel Approaches To Global Benchmarking Of Risk-Adjusted Surgical Outcomes

The above proposal has been reviewed by the Department of Surgery Research Committee. I am pleased to inform you that the committee approved the scientific merit of the study, and endorse the protocol for submission to the relevant ethics committee.

Please use the above project number in all future correspondence.

Yours sincerely

A handwritten signature in black ink, appearing to read 'T Pennel'.

DR TIMOTHY PENNEL
CHAIRMAN: RESEARCH COMMITTEE

REVIEWERS COMMENTS

From: worldjsurg@ohsu.edu

To: spnric004@myuct.ac.za, rts25@mgh.harvard.edu

CC:

Subject: World Journal of Surgery - Decision on Manuscript ID WJS-17-02-0186 (D-RW-03)

Body: 08-Apr-2017

Dear Dr. Spence:

Manuscript ID WJS-17-02-0186 entitled "Derivation, Validation and Application of a Low to middle Income Countries Specific Index to Predict Outcome Following Noncardiac Surgery" which you submitted to the World Journal of Surgery, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

The reviewers showed considerable interest in your manuscript but do not feel that it is acceptable for publication in its present form. We would be ready to consider a revised version if you are able to completely and satisfactorily answer their questions and revise the manuscript to address each point. The editors of the World Journal of Surgery reserve the right to reject a future revision of this manuscript if you fail to adequately address the reviewer's comments.

To revise your manuscript, log into <https://mc.manuscriptcentral.com/wjs> and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Please also highlight the changes to your manuscript within the document by using the track changes mode in MS Word or by using bold or colored text. Once the revised manuscript is prepared, you can upload it and submit it through your Author Center.

When submitting your revised manuscript, please respond to the comments made by the reviewer(s) in the space provided. You can use this space to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

Because we are trying to facilitate timely publication of manuscripts submitted to the World Journal of Surgery, your revised manuscript should be uploaded as soon as possible. If it is not possible for you to submit your revision in a reasonable amount of time, we may have to consider your paper as a new submission.

Once again, thank you for submitting your manuscript to the World Journal of Surgery and I look forward to receiving your revision.

With kind regards,
Dr. John Hunter
Editor-in-Chief, World Journal of Surgery
hunterj@ohsu.edu

Reviewer(s)' Comments to Author:

Reviewer: 1

Comments to the Author
Dear Colleagues

Thank you for the opportunity to review the paper; 'Derivation, Validation and Application of a Low to middle Income Countries Specific Index to Predict Outcome Following Noncardiac Surgery'. This is a paper describing the development of a simple, feasible and pragmatic risk stratification tool, which is proposed for noncardiac surgical patients in low and middle income countries.

General comments:

1. I particularly like the approach of only including risk factors which are feasible to collect i.e. >90% compliance. This is a strength of this approach to development of a preoperative risk model.
2. It is claimed that this is a model for low to middle income countries (LMIC). The model was however developed in an upper middle income country (South Africa), and hence it may not be as robust in a low-income country (LIC). This should be acknowledged in the limitations section in the discussion. The risk factor which particularly concerns me in a low-income country would be the preoperative transfusion, which may not always be possible in a very resource limited environment. It is possible another risk factor which may reflect the need for preoperative transfusion may need to be substituted in a low-income country model. All references to LMIC should be replaced by MIC. In the discussion, it should be suggested that the model may have applicability in LIC, but this would need to be tested and externally validated.
3. This is a model for general and vascular surgery only. This should be reflected in the title and the abstract. To claim, that this is a model for noncardiac surgery, it is important to externally validate the model in a broader noncardiac surgery population. Indeed, this would also need consideration of other non-vascular, non-general surgical procedures, which may need to be added to the model due to their impact on perioperative outcomes.

Specific comments:

1. Title and abstract:

- a. Should only claim MIC, and general and vascular surgery.

2. Introduction:

- a. Please add preoperative to 'application'; 'The Surgical Apgar Score has been well validated globally but is based on data collected intraoperatively at 5-minute intervals, limiting its application', as your paper is about a pragmatic preoperative score.

3. Methods:

- a. The fact that the data collection was limited to general surgical and vascular surgical patients, does potentially limit the applicability of the score. This should also be acknowledged in a limitations section in the discussion.
- b. The data collection tool from the ACS website, should be added as web supplementary material for this paper, as should the REDCap data collection tool. This will help readers/ investigators who may want to externally validate this score in the future, in their own surgical settings.
- c. 'postoperative' should not be hyphenated.
- d. Step 2. The in-hospital mortality was 24 patients and for morbidity it was 127 for South Africa. Step two is unclear to me. Did you only include risk factors with a $p < 0.1$ for SA mortality, in developing the model for the combined outcome of mortality and any morbidity ($n=127$)? If this is the case ideally one would not want to include more than about 13 variables in the logistic regression (events per variable (EPV) of 10).(1) You have 23 variables in Table 1, and 23 variables in the full model in table 3. Please make it clear to the readers which variables were entered into the logistic regression model, in what order they were entered and why that particular order of entry, so that the reader can understand the EPV in your multivariate analysis. At present, it appears that following the entry of the surgical procedure, various combinations of entries were conducted, which will result in an overfitted model.(2)
- e. Step 3. It is unclear to me, whether the derivation and validation cohort of the NSQIP database only included surgical procedures considered general surgical and/ or vascular procedures, or did the NSQIP cohort also include other non vascular, non general surgical procedures? It is unclear if the 'thoracic' procedures, where 'thoracic' procedures performed routinely by 'general surgeons' or 'thoracic surgeons'. This needs to be clarified.
- f. Step 3. I would prefer the term 'ROC curve' as opposed to 'ROC' alone.
- g. Step 3. When discussing the performance of the ROC curve, it is not the ROC, but rather the AUC (area under the curve), which is presented as the c-statistic. So it would be better to read 'highest c-statistic' as opposed to 'highest ROC'. This is seen a few times in varying places in the text. These all need to be corrected e.g. had an ROC of 0.8672, should read; had a c-statistic of 0.8672'.

h. Step 3. Spelling 'individually'

i. Step 4. Please clarify if 'any morbidity' includes mortality.

j. Step 4. For future reference the 'Hosmer-Lemeshow test' for calibration is not ideal for assessing calibration.² It is most likely to be of very limited value in your South African cohort with a small sample size.

4. Results:

a. The following statement is problematic, and I cannot agree with it; 'patients in the SA dataset appeared to be sicker with a higher proportion of emergency cases, diabetics, symptomatic dyspnea, functional dependency, congestive heart failure, chronic obstructive pulmonary disease, acute renal failure, disseminated cancer and preoperative sepsis'.

Firstly, it is well documented that significantly more South African patients (3) have a lower ASA score than patients in high-income countries,⁴ which is consistent with your data. Furthermore, ASA is robust predictor of perioperative outcome⁽⁴⁾ despite some subjectivity to its assessment. (There are techniques/ mechanisms to improve the reliability of the documentation of the ASA score, but I believe this is beyond the scope of your paper, and will detract from the more important aspect of your risk prediction model).

Secondly, you may be mixing other risk factors which do not necessarily change the ASA score e.g. urgent and emergent surgery (which is significantly more common in your cohort than the NSQIP cohort). Urgent and emergent surgery may result in more symptomatic patients from the surgical pathology e.g. associated dyspnoea, acute renal failure and sepsis.

Thirdly, the absolute number of patients with some of the pathologies you describe, are still far smaller than the number of patients who are ASA 3 or above. What you are essentially describing is a result of i) a single tertiary centre compared to data from a range of levels of care, and ii) a health system which is characterised by late surgical presentation,⁽³⁾ compared to a more functional high-income country health system with early presentation of surgical pathologies.

b. There is a selection bias in the South African cohort which has resulted in the higher proportion of certain types of operative procedures. This is a function of a single, tertiary centre as the only site for data capture.

c. ICU admission. This is a post hoc analysis of the Codman Score associated with ICU admission. I have reservations of its applicability, as it reflects the admissions to ICU at a single hospital, which may be variously affected by bed availability, prognosis of patient, etc. I would remove this from the results. If one, wanted to know the use of the Codman Score for predicting ICU admission, it would be better to study this prospectively, where the need for ICU was determined (irrespective of whether the patient received ICU admission), and then this is correlated with the Codman Score.

5. Discussion:

a. I would suggest rewriting the discussion as suggested by Docherty and Smith,⁽⁵⁾ as at present I believe the discussion does not explicitly address the strengths, limitations and implications of the research, and as such the emphasis of the discussion is too focused on a small part of the entire research project i.e. in particular O/E ratio, and ASA score, which does not do justice to the important research presented in the paper. Based on Docherty and Smith's proposal this is what I would suggest;

i. Statement of principal findings: Feasible, 7 variable model for general and vascular surgery in a MIC.

ii. Strengths and weaknesses of the study: Strengths; feasible, and simple. Weaknesses, single tertiary centre, limited to general and vascular surgery from a MIC, and not a LIC. Statistical weaknesses, possibly overfitted model.

iii. Strengths and weaknesses in relation to other studies, discussing particularly any differences in results: Strengths, simple parsimonious model when compared to NSQIP.

iv. Meaning of the study, possible mechanisms and implications for clinicians or policymakers: If the model could be externally validated, it is simple enough that an expected outcome for a country could be obtained of a period of time, after which the O/E ratio, may be useful for benchmarking performance within a country.

v. Unanswered questions and future research: This is a model which may be feasible in LIC. It is necessary however, that there is external validation of the model in LIC. Furthermore, to extend the model beyond general and vascular surgery, it will need possible inclusion of other surgical procedures into the model.

b. Other comments; Remove LMIC and replace with MIC.

c. Statement; 'In contrast, our findings question the reliability of ASA rating in such diverse settings.' I cannot agree with this from a small single centre study, particularly when 28 countries of over 46 000 patients have shown its reliability,⁴ and 50 centres in South Africa have shown its reliability.⁽²⁾ This is discussed above, and should be removed. I believe the inappropriate focus on the ASA score is detracting from the more important contribution your model could make to surgical outcomes, and in

all likelihood, the ASA score will always be one of the included risk factors in your model.

6. References:

a. Please complete reference 13.

7. Tables

a. Table 2. Please clarify if 'any morbidity' includes mortality

b. Table 3. I would remove 'model with 23 variables'. This is an overfitted model, based on your sample size, and hence should not be included in the table.

8. Figures

a. Figure 3.

i. Spelling 'cumulative'

ii. CI for the outcomes would be desirable. I expect that there is a lot of overlap between scores. It may be more advantageous to have two or three score categories, where there is little overlap of CI for outcomes.

Thank you for the opportunity to review this paper.

Regards

Bruce Biccard

Conflict of interest: I am an author and co-investigator in the SASOS paper which I reference in some of my discussions of this manuscript above.

References

1. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49(12):1373-9. doi: S0895-4356(96)00236-3 [pii] [published Online First: 1996/12/01]
2. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 2014;35(29):1925-31. doi: 10.1093/eurheartj/ehu207
3. Biccard BM, Madiba TE, South African Surgical Outcomes Study I. The South African Surgical Outcomes Study: A 7-day prospective observational cohort study. *S Afr Med J* 2015;105(6):465-75. doi: 10.7196/SAMJ.9435
4. Pearse RM, Moreno RP, Bauer P, et al. Mortality after surgery in Europe: a 7 day cohort study. *Lancet* 2012;380(9847):1059-65. doi: 10.1016/S0140-6736(12)61148-9 [published Online First: 2012/09/25]
5. Docherty M, Smith R. The case for structuring the discussion of scientific papers : Much the same as that for structuring abstracts *BMJ : British Medical Journal* 1999;318(7193):1224-25.

Reviewer: 2

Comments to the Author

This is an important manuscript that moves forward the field of global surgery. A hearty congratulations to the team for taking on this task, and cleverly so. The Lancet Commission championed the need for benchmarks and standards -- for indicators of the strength of the surgical system and was met with widespread acceptance on the indisputable assertion. It did not, however, prescribe a specific approach to how make sense of indicators collected in widely different health systems with different patient populations. The work by Spence and team moves us closer to understanding how to measure health system quality and strength. It also provides a valuable tool for surgeons operating in the trenches-- perhaps for the first time ever -- to benchmark their outcomes against their peers in other parts of the world. Clearly, the Codman score, as described, will still has many years of work ahead to tweak and revise based on its performance in different environments, and perhaps even further simplification and less reliance on subjective assessments like ASA. Nonetheless, it provides a valuable first step based on existing and accepted measurements systems for quality, consistent with calls from works before it (e.g. The Lancet Commission), and simplifies and adjusts it to being able to be collected in the low-resource environment.

My one point of contention with the framing of the manuscript, however, is that this really has to do with simplifying NSQIP to be able to be collected with fewer variables than the data and time-intensive manner it currently requires. It could be used in Boston just as well as in Cape Town. And while it is true that the data simplification makes it more reasonable to be used in an LMIC, it is not actually

developed or tested in a true "low-resource" environment. Instead the score is developed and tested in a South African hospital that, true, is in an "LMIC" (a horrible catch-all phrase to begin with) but wouldn't fit my definition of a 'low-resource' setting.

The authors peripherally discuss this in the limitations section, and I understand the reality of the logistical/financial limitations of testing this in multiple environments around the world. But the absence of testing in some of the lowest-resource environments presents real challenges to the Codman score's applicability to these settings.

For example, usage of the blood transfusion requirement of 4 units or greater in the score predicates that we are operating in environments where blood transfusion/blood banking is an option. This limits the applicability of this score away from broad swaths of the world. This is unfortunate, because the rest of the score is easily collectible. Can the authors comment on whether a modification can be introduced that is based on a starting Hemoglobin (this is also a challenge, granted, but at least provides an objective measure and is more broadly available in low-resource hospitals in rural settings than a the scarce resource of blood).

Date Sent: 08-Apr-2017

From: worldjsurg@ohsu.edu
To: spnric004@myuct.ac.za, rts25@mgh.harvard.edu
CC:
Subject: World Journal of Surgery - Decision on Manuscript ID WJS-17-02-0186.R1 (D-RW-02)
Body: 18-Jun-2017

Dear Dr. Spence:

Manuscript ID WJS-17-02-0186.R1 entitled "Derivation, Validation and Application of A Pragmatic Risk Prediction Index for Benchmarking of Surgical Outcomes" which you submitted to the World Journal of Surgery, has been reviewed. The comments of the reviewer(s) are included at the bottom of this letter.

The reviewer(s) suggest some minor revisions to your manuscript. Therefore, I invite you to respond to the reviewer(s)' comments and revise your manuscript.

To revise your manuscript, log into <https://mc.manuscriptcentral.com/wjs> and enter your Author Center, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, revise your manuscript using a word processing program and save it on your computer. Please also highlight the changes to your manuscript within the document by using the track changes mode in MS Word or by using bold or colored text.

Once the revised manuscript is prepared, you can upload it and submit it through your Author Center.

When submitting your revised manuscript, please respond to the comments made by the reviewer(s) in the space provided. You can use this space to document any changes you make to the original manuscript. In order to expedite the processing of the revised manuscript, please be as specific as possible in your response to the reviewer(s).

IMPORTANT: Your original files are available to you when you upload your revised manuscript. Please delete any redundant files before completing the submission.

Because we are trying to facilitate timely publication of manuscripts submitted to the World Journal of Surgery, your revised manuscript should be uploaded as soon as possible. If it is not possible for you to submit your revision in a reasonable amount of time, we may have to consider your paper as a new submission.

Once again, thank you for submitting your manuscript to the World Journal of Surgery and I look forward to receiving your revision.

With kind regards,
Dr. John Hunter
Editor-in-Chief, World Journal of Surgery
hunterj@ohsu.edu

Reviewer(s)' Comments to Author:
Reviewer: 1

Comments to the Author
Dear Colleagues

Thank you for your revisions and explanations.

I have a few minor revisions; (I have used the review pdf page and line numbers).

1. Page 8, line 13: Remove text 'The tertiary outcome collected in the SA dataset was post-operative Intensive Care Unit (ICU) admission.' as this is now redundant.

From: worldjsurg@ohsu.edu
To: spnrnc004@myuct.ac.za, rts25@mgh.harvard.edu
CC: Yvonne.Chan@springer.com
Subject: World Journal of Surgery - Decision on Manuscript ID WJS-17-02-0186.R2 (D-RW-01)
Body: 07-Jul-2017
Thank you and congratulations.
Dear Dr. Spence:
Regards
Your manuscript entitled "Derivation, Validation and Application of A Pragmatic Risk Prediction Index for Benchmarking of Surgical Outcomes" has been accepted in its current form for publication in the World Journal of Surgery contingent upon a Crosscheck scan. After acceptance, all WJS material is put through a computer search comparing the new material with all previously published scientific work. We will contact you if any significant "overlap" is demonstrated.

Thank you for your fine contribution. On behalf of the Editors of the World Journal of Surgery, we look forward to your continued contributions to the Journal.

With kindest regards,
Dr. John Hunter
Editor-in-Chief, World Journal of Surgery
hunterj@ohsu.edu

Date Sent: 07-Jul-2017

From: worldjsurg@ohsu.edu

To: spnrnc004@myuct.ac.za, rts25@mgh.harvard.edu

CC: Yvonne.Chan@springer.com

Subject: World Journal of Surgery - Decision on Manuscript ID WJS-17-02-0186.R2 (D-RW-01)

Body: 07-Jul-2017

Dear Dr. Spence:

Your manuscript entitled "Derivation, Validation and Application of A Pragmatic Risk Prediction Index for Benchmarking of Surgical Outcomes" has been accepted in its current form for publication in the World Journal of Surgery contingent upon a Crosscheck scan. After acceptance, all WJS material is put through a computer search comparing the new material with all previously published scientific work. We will contact you if any significant "overlap" is demonstrated.

Thank you for your fine contribution. On behalf of the Editors of the World Journal of Surgery, we look forward to your continued contributions to the Journal.

With kindest regards,
Dr. John Hunter
Editor-in-Chief, World Journal of Surgery
hunterj@ohsu.edu

Date Sent: 07-Jul-2017