Title: Recalibration And External Validation Of The Risk Analysis Index: A Surgical Frailty Assessment Tool

Authors:

Shipra Arya, MD, SM[1]

Patrick Varley, MD[2]

Ada Youk, PhD[3,4]

Jeffrey Borrebach, MS[5]

Sebastian Perez MS[6]

Nader N Massarweh MD MPH[7]

Jason M. Johanning, MD, MS[8]

Daniel E. Hall, MD, MDiv, MHSc[2,3,5]

Affiliations

[1]Division of Vascular Surgery, Stanford University School of Medicine, Stanford, CA and Surgical Service Line, VA Palo Alto Healthcare System, Palo Alto, CA

[2]Department of Surgery, University of Pittsburgh, Pittsburgh PA

[3]Center for Health Equity Research and Promotion, VA Pittsburgh Healthcare System, Pittsburgh, PA

[4]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

[5]Wolff Center at UPMC, University of Pittsburgh Medical Center, Pittsburgh, PA

[5]Department of Surgery, Emory University, Atlanta GA

[7]Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey VA Medical Center; Michael E DeBakey Department of Surgery, Baylor College of Medicine, Houston, TX

[8]Department of Surgery, University of Nebraska Medical Center and Nebraska Western Iowa VA Health System, Omaha NE

Corresponding Author:

Daniel E Hall, MD

UPMC Presbyterian, Suite 1264

200 Lothrop St

Pittsburgh, PA 15213

412.647.0421

hallde@upmc.edu

Brief Title: Recalibration and external validation of the Risk Analysis Index

Word Count: 3944, not including tables and figures, abstract, references

Tweet: An improved RAI frailty assessment tool for screening surgical patients.

Conflict of Interest Disclosures:. Dr. Johanning holds intellectual property on frailty through FutureAssure, LLC. No other disclosures are reported.

Funding/Support: This research was supported by the US Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development (I21 HX-002345 and XVA 72-909 [Hall], CIN 13-413 [Massarweh]) and NIH/NIA 5R03AG050930 (Arya).

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Disclaimer: The opinions expressed here are those of the authors and do not necessarily reflect the position of the Department of Veterans Affairs or the US government.

MINI-ABSTRACT

This study recalibrates the previously developed Risk Analysis Index (RAI) in a large Veteran surgical registry and externally validates it using a national surgical cohort and a survey instrument. The generalizability of the RAI across various surgical populations makes it an ideal instrument for frailty assessment in preoperative patients.

STRUCTURED ABSTRACT

Objective and Background: The Risk Analysis Index (RAI) predicts 30, 180 and 365-day mortality based on variables constitutive of frailty. Initially validated, in a single-center Veteran hospital, we sought to improve model performance by recalibrating the RAI in a large, Veteran surgical registry, and to externally validate it in both a national surgical registry and a cohort of surgical patients for whom RAI was measured prospectively before surgery.

Methods: The RAI was recalibrated among development and confirmation samples within the Veterans Affairs Surgical Quality Improvement Program (VASQIP; 2010-2014; N=480,731) including major, elective non cardiac surgery patients to create the revised RAI (RAI-rev), comparing discrimination and calibration. The model was tested externally in the American College of Surgeons National Surgical Quality Improvement Program dataset (NSQIP; 2005-2014; N=1,391,785), and in a prospectively collected cohort from the Nebraska Western Iowa Health Care System VA (NWIHCS; N=6,856).

Results: Recalibrating the RAI significantly improved discrimination for 30-day [c=0.84 to 0.86], 180-day [c=0.81 to 0.84] and 365-day mortality [c=0.78 to 0.82](p<0.001 for all) in VASQIP. The RAI-rev also had markedly better calibration (median absolute difference between observed and predicted 180-day mortality: decreased from 8.45% to 1.23%). RAI-rev was highly predictive of 30-day mortality (c=0.87) in external validation with excellent calibration (median absolute difference between observed and predicted 30-day mortality: 0.6%). The discrimination was highly robust in men (c=0.85) and women (c=0.89). Discrimination also improved in the prospectively measured cohort from NWIHCS for 180-day mortality [c=0.77 to 0.80] (p<0.001).

Conclusions: The RAI-rev has improved discrimination and calibration as a frailty screening tool in surgical patients. It has robust external validity in men and women across a wide range of surgical settings and available for immediate implementation for risk assessment and counseling in preoperative patients.

INTRODUCTION

Patients over 65 years of age undergo almost one-third of surgical procedures in the US, presenting a unique set of challenges for surgeons, patients and their families[1,2]. Studies have shown the need to look beyond morbidity and mortality at 30 days and focus on patient centered outcomes including preserving function, maintaining independence and avoiding readmissions and institutionalization for older patients[3,4]. It is imperative that methods of surgical risk-assessment in this population be continuously improved to identify patients vulnerable to adverse outcomes. Frailty is a syndrome of physiological decline that places patients at increased risk for death and disability[5-7]. Originally identified by geriatricians in community-dwelling adults, the concept of frailty has been successfully applied to surgical populations to identify those at risk for poor outcomes[8-10]. The concept of frailty not only helps identify high-risk patients, but more importantly provides a framework for placing the proposed surgical intervention into the context of the patient's overall health.

Several tools for measuring surgical frailty have been proposed, including the modified Frailty Index (mFI)[11], the Fried frailty phenotype[5] and a complex multi-modal assessment developed by Robinson[3]. However, none of these tools are suitable for real-time screening of large populations. For example, the approaches developed by Fried and Robinson require specialized equipment and labor-intensive assessment of physical performance, making them ideal for research protocols, but impractical for system-wide screening. The mFI has been validated in American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP) registries, but it has never been developed or validated as a prospective survey instrument, and now that 6 of the 11 required variables have been phased out of ACS NSQIP, it

is obsolete[12]. As such, there remains the need for a pragmatic frailty assessment suitable for screening.

Our prior work aimed to meet this need of systemwide screening, by developing and implementing the Risk Analysis Index (RAI)[13,14]. The RAI is a tool based on the accumulation of deficits model of frailty derived from the Minimum Data Set Mortality Risk Index-Revised (MMRI-R) instrument[15], comprising 14 variables including age, sex, weight loss, poor appetite, congestive heart failure, dyspnea, renal failure, presence of disseminated cancer, functional status, cognitive decline and living status. It has been validated in two forms, the "administrative RAI" (RAI-A) calculated from variables contained in VA Surgical Quality Improvement Project (VASQIP) or ACS-NSQIP and the "clinical RAI" (RAI-C) calculated from responses to a survey instrument with significant correlation between the two forms (r=0.48)[13]. More importantly, our initial quality improvement experience with this tool showed that implementation of routine frailty screening with the RAI as part of a Frailty Screening Initiative (FSI) was associated with reduced institutional surgical mortality[14].

Given the compelling initial experience with the RAI, we sought to examine whether the calibration and discrimination of the RAI could be improved and generalized to non-veteran surgical patients, especially women who constitute a minority of Veterans. Because the initial MMRI-R was developed in a non-surgical population, we hypothesized that re-weighting the RAI score specifically for surgical patients would result in improved predictive performance. To accomplish this, we recalibrated the RAI scoring paradigm utilizing the VASQIP database. We then externally validated the revised RAI (RAI-rev) in elective surgery patients in the ACS NSQIP database with age and gender distributions representative of US surgical patients[16]. We also applied the revised scoring paradigm to the survey instrument version of the tool (RAI-C-

rev), testing it in the original cohort of patients from the Nebraska Western Iowa Health Care System (NWIHCS) VA hospital in which the RAI was measured preoperatively.

METHODS

Data Sources and Patient Selection

Established in 1991, VASQIP rigorously obtains information regarding surgical procedures from all VA hospitals in order to facilitate surgical quality improvement efforts. Complete descriptions of this dataset have been previously published.[17] We chose VASQIP because, unlike ACS NSQIP, it includes mortality data beyond 30-days permitting calibration to 180-day mortality. After the Institutional Review Board (IRB) of the VA Pittsburgh Healthcare System determined this study to be exempt, we obtained VASQIP records for all available CPT codes linked to 365-day mortality data for non-cardiac surgical cases occurring between April 1, 2010 and March 31, 2014.

For the purpose of external validation, we obtained the ACS NSQIP participant user file datasets from 2005-2014. Detailed descriptions of this dataset and its methodology are published elsewhere[18], and the study was deemed exempt by the Emory University and Stanford University IRBs. Within ACS NSQIP, we identified elective, non-cardiac surgical patients with complete case information on the RAI variables. The initial validation of the RAI-C was conducted at the NWIHCS where it was administered preoperatively to 6856 unique patients between July 2011 and September 2015.[13] Pursuant to approvals from the IRBs at the NWIHCS (ID#01080) and VA Pittsburgh Health Care System (Pro#1666), de-identified copies of the original dataset were obtained to validate the revised scoring paradigm.

Calculation of the RAI

Using VASQIP, the RAI-A was calculated for all patients using methods described in previous work[13]. We then built logistic regression models using RAI-A to predict mortality at 30-, 180-, and 365 days, generating c-statistics as a measure of discrimination and Aikake Information Criterion (AIC)[19] and Maximum R^2 (Max R^2)[20] statistics as measures of calibration. To further assess calibration, we calculated the absolute value of the difference between the observed and predicted mortality for each integer value of RAI-A, reporting the median difference and the interquartile range. Finally, across all integer values of RAI-A, we reported the proportion of cases where the 95% confidence interval for the observed mortality included the predicted mortality, calling this statistic "overlap" as described elsewhere[21].

Recalibration of the RAI-A for a Surgical Population

We randomly split the VASQIP data equally into development and confirmation samples. In the development sample, we built a new logistic regression model using the RAI variables to predict 180-day mortality, controlling for the hospital site. We then tested the development model parameters in the confirmation sample, fixing the parameters as derived and calculating a Receiving Operator Characteristic (ROC). Confirmation was defined a priori as no statistical difference between the c-statistics using the nonparametric methods described by Delong, et al [21]. We then used the model parameters to define a new scoring system for the revised RAI-A, hereafter denoted as RAI-rev. We then applied the RAI-rev to calculate a frailty score for each record and (using logistic regression) to predict mortality at 30, 180 and 365 days post-surgery. Similar statistics as described above were used to ascertain discrimination and calibration. Finally, for each integer value of RAI-rev, we calculated sensitivity, specificity, positive and negative predictive values (PPV & NPV), and predicted and observed 30- and 180-day mortality.

External Validation of RAI-A: ACS NSQIP Data and Methods

We used methods identical to those described for the VASQIP data to calculate the RAI-rev and to predict 30-day mortality. To visually display changes in calibration we plotted the predicted mortality across the range of RAI scores along with the observed mortality for each integer value of RAI with exact confidence intervals. In addition to the overall cohort, men and women were plotted separately.

Validation of RAI-C: NWIHCS data and methods

We applied the revised scoring paradigm to the RAI-C to assess model improvement. The scoring paradigm for the RAI-C is identical to the RAI-A except that the RAI-C measures ADLs with increased granularity. VASQIP and ACS-NSQIP code only 3 levels of physical function (i.e., independent, partially dependent and totally dependent). In contrast, the RAI-C survey uses a 5-point Likert scale to assess 4 domains of ADLs (i.e., mobility, eating, toileting and personal hygiene) to generate a combined ADL score ranging from 0 to 16[13]. In order to preserve the same range of scores for patients with and without cognitive decline, we scaled the ADL*Cognitive Decline score as shown in Table 3. Methods identical to those described above were used to compute the revised RAI-C [RAI-C-rev] and model mortality at 30 and 180 days.

All analyses were completed with STATA (StataCorp. 2015.Statistical Software: Release 14. College Station, TX: StataCorp LP) except ACS NSQIP analyses (completed with SAS version 9.2).

RESULTS

After removing missing or out of range values, the VASQIP, ACS-NSQIP and NWIHCS data sets contained 480,731, 1,391,785 and 6856 records, respectively. Demographic and

clinical characteristics of cohorts as well as the components of the RAI are detailed in Table 1. The VASQIP cohort was 92.2% male whereas the ACS-NSQIP cohort was 58.0% female. 30-day mortality were similar: 1.1% in VASQIP and 1.0% in ACS-NSQIP, while lower in the NWIHCS data 0.4%.

Testing model performance of RAI-A in VASQIP

The RAI-A score computed according to the original parameters demonstrated model discrimination similar to that seen in the original sample of Veteran patients drawn from the NWIHCS: c-statistics for 30-, 180-, and 365-day mortality were 0.842 (95% CI 0.835-0.848), 0.813 (95% CI 0.810-0.817), and 0.784 (95% CI 0.781-0.787) respectively (Table 2).

Recalibration of the RAI-A

There was no statistically significant difference between the c-statistics for the RAI-rev in the development and confirmation samples (0.847 vs. 0.848, respectively, p=0.718). To allow clinical application of the RAI-rev, point estimates for each variable were scaled to an integer value representing the points assigned for each element, yielding a raw score ranging from 0-128. However, because the original RAI-A ranged from 0-81, the RAI-rev was rescaled to the same range in order to facilitate direct comparisons (Table 3). The RAI-rev treats age as a continuous variable (interval in years) to reflect better the wider range of age in this sample of surgical patients.

Comparing the RAI-A and Recalibrated RAI-A (RAI-rev)

Compared to the original RAI-A, the RAI-rev demonstrates statistically significantly improved discrimination and calibration for mortality (Table 2, Figure 1). For example, at 180-days the discrimination improves from 0.813 to 0.842 ($p<0.001$), the Max R^2 increases from 0.211 to 0.255, and the AIC decreases from 120,967.0 to 114,881.8. The improvement in

calibration is most apparent when comparing the differences between the observed and predicted mortality across the range of original and revised RAI scores (Figures 1a & 1b). The median absolute difference between observed and predicted mortality fell from 8.45% (IQR 2.48-17.16) to 1.23% (IQR 0.12-8.5) and the proportion of records where the 95% CI of the observed mortality overlapped the predicted mortality increased from 22.1% to 46.5%. Taken together these data demonstrate that across all VASQIP-eligible surgical procedures, performance of the RAI-rev is significantly improved relative to the older RAI-A.

External Validation of RAI-rev: ACS-NSQIP Data

Discrimination for the RAI-rev for 30-day mortality was excellent (c=0.870, 95% CI 0.867-0.873) with a Max R^2 of 0.222 and an AIC of 118,997.0 (Table 2). In addition, the median absolute difference between observed and predicted mortality was only 0.6% (IQR 0.04-10.9) and the proportion of records where the 95% CI of the observed mortality overlapped the predicted mortality was 41.7%. The robust calibration of RAI-rev in the ACS-NSQIP data are most easily appreciated graphically (Figure 1c). Agreement between observed and predicted mortality was better for lower RAI values as compared to higher ones. The wide confidence intervals in the upper range of RAI score reflects the small number of patients with these scores (Figures 1a, 1b, and 1c).

External Validation of RAI-rev in Women

When women and men were modeled separately, the recalibrated RAI-rev demonstrated excellent discrimination and calibration (Table 2 and Figure 2). For example, c-statistics for 30-day mortality were 0.885 for women and 0.845 for men (Table 2). Of note, the proportion of records where the 95% CI for observed mortality overlapped predicted mortality was better in women (47.2%) as compared to men (44.4%). Calibration plots for the RAI-rev had similar

findings among men and women in terms of better agreement of observed and predicted mortality at lower RAI scores (Figure 2).

Validation of RAI-C: NWIHCS Data

We accurately recapitulated the previously published c-statistics and compared them to the revised RAI-C model estimates. As expected, discrimination for mortality at 30-days and 180-days improved (Table 2), but statistically significantly for only 180-day mortality from 0.772 to 0.804 ($p<0.001$). As in the other samples, calibration improved dramatically with the revised RAI-C (Table 2, and Figure 1d) and the median absolute difference between observed and predicted 180-day mortality fell from 0.9% (IQR 0.4-1.5) to 0.3% (IQR 0.2-1.1).

Choosing thresholds for RAI-rev: Sensitivity, Specificity, PPV, NPV, Predicted Mortality

Given that the RAI-rev is meant to identify patients at increased risk for postoperative morbidity and mortality, Table 4 reports relevant predictive parameters for selected threshold values of RAI-rev in the VASQIP and ACS-NSQIP cohorts and RAI-C-rev in the NWIHCS cohort (See eTables 1, 2 and 3 for similar parameters reported for each integer value of RAI-rev and RAI-C-rev). We found that in both registry cohorts, an RAI-rev score of 25 indicates a predicted mortality approximately equal the overall observed mean mortality (~1%), and that for each subsequent 5-point rise in RAI-rev, the predicted mortality approximately doubles. For example, in the ACS-NSQIP cohort, the predicted 30-day mortality for RAI-rev scores of 25, 30, 35, 40 and 45 are 1.1%, 2.3%, 4.9%, 10.2% and 20.1 %, respectively. Similar rates of doubling are observed in the VASQIP cohort for both 30 and 180 day mortality (Table 4). Of note, most patients scored at or below RAI-rev scores of 25 (e.g., 80.0% and 74.5% in the ACS-NSQIP and VASQIP cohorts, respectively). A threshold of RAI-rev ≥ 30 approximates the highest risk decile of patients with a predicted mortality risk at least twice mean mortality risk of the entire

cohort. At this threshold, sensitivity is only 58-59%, but the negative predictive value (NPV) is 99.6% and 98.4% for the ACS-NSQIP and VASQIP cohorts, respectively. Thus, a cutoff of RAI-rev <30 is extremely effective at identifying low risk patients, while further geriatric testing may be needed for patients with RAI-rev ≥30 to diagnose specific risks (e.g. sarcopenia, weakness or cognitive decline).

Similar relationships are observed for the RAI-C-rev, although the thresholds are somewhat higher because, as described previously[13], the survey mode of administration is more flexible and open to interpretation than the strict SQIP coding rules, yielding higher RAI-C scores for a given mortality risk. Thus, in the NWIHCS data a threshold of RAI-C-rev ≥ 37 identifies 14.0% of the cohort as frail with a predicted 180-day postoperative mortality of 4.3% which is approximately twice the 1.8% overall mean mortality in the cohort. Mortality approximately doubles with each subsequent 8-point rise in RAI-C-rev to 10.3%, and 22.4% (Table 4).

DISCUSSION

With surgeons facing a "silver tsunami" of older patients in the United States, surgical and geriatric societies have recognized the importance of pragmatic tools for accurately quantifying overall risk in the context of shared discussions regarding goals of care[22,23]. Our current study advances our prior work regarding the RAI in several important ways. First, this study strengthens the RAI by recalibrating it in a large, national sample of Veteran patients. We found the RAI-rev has significantly improved model discrimination (c=0.84) and calibration. Second, our current study demonstrates robust external validity of the RAI in a nationally representative, non-veteran surgical registry (ACS-NSQIP) with excellent discrimination in

predicting 30-day mortality (c= 0.87). Third, we have shown that the RAI-rev performs equally well in women (c=0.89) and men (c=0.85), making it generalizable for all surgical patients. Finally, we show that the revised scoring paradigm improves the performance of the revised RAI-C, a survey instrument version of the tool that takes less than 2 minutes to administer and is suitable for point-of-care risk assessment and real-time counseling in preoperative patients.

These improvements place the revised RAI on par with some of the best tools for predicting short and long-term mortality. For example, the Hospital-patient One-year Mortality Risk (HOMR) model uses diagnosis codes from a previous hospitalization to predict 365-day mortality. It has been validated in cohorts from Ontario, Alberta and Boston with c-statistics ranging from 0.89 (95% CI 0.87-0.91) to 0.92 (95% CI 0.91 to 0.92)[21]. The RAI-rev performance is similar to HOMR, is calibrated specifically to a surgical population, uses 20 fewer variables and can be administered prospectively to guide real-time clinical decisions[14]. In addition, we show high precision, as RAI-rev predictions are within 1% of observed mortality ranging for a wide range of frailty scores from zero up to ~45. Precision falters for the highest RAI-rev scores (e.g., RAI-rev > 45), but these records account for only 1-3% of the cohort in both registry datasets. The predicted 30-day mortality and 6-month mortality is closer to 20% and 40% respectively for RAI-rev scores of 45 (Figure 1b). The survey instrument version also has a 6 month mortality of around 22% for RAI-C-rev of 53 or greater. These robust and sobering mortality predictions are consistently above a threshold that would give most surgeons pause, and which many patients would find unacceptable.

The RAI is the only frailty assessment tool now validated in multiple surgical populations and shown to be associated with improved survival in clinical practice through systemwide screening of elective surgery patients[14]. The tool itself gives higher weight to male gender as part

of its scoring scheme. However, our study shows it performs equally well in predicting postoperative mortality in women, making it ideal for adoption in any clinical practice setting. It is also the only index of surgical frailty that has been developed for prospective assessment through a validated survey instrument that takes less than a minute to administer and has proven feasible for real-time screening in a variety of contexts across the country (e.g., the RAI-C). Recent epidemiological research recommends recalibration of frailty scores in populations outside which they were developed before clinical use[24], and we have successfully done that in this analysis.

The revised RAI has several advantages over the other widely-used measure of surgical frailty in large datasets (e.g., the modified Frailty Index or mFI)[11]. Both are based on the accumulated deficits model of frailty developed by Rockwood, et al[6], but the RAI is a weighted model as we don't believe all deficits are equal. Furthermore, the RAI assesses deficits across five domains of frailty [physical (comorbidity), functional, social, nutritional and cognitive], thus ensuring that it is a more comprehensive measure than the mFI which has been criticized as merely a uni-dimensional index of multimorbidity similar to a Charlson Score[25]. Furthermore, unlike the RAI, the mFI has never been deployed or validated prospectively—all published mFI data rely on a single, retrospectively collected administrative dataset, namely the ACS-NSQIP. Finally, due to changes in ACS-NSQIP data capturing between 2012-2015, the mFI cannot be calculated anymore because half of the necessary variables were phased out of the program[12]. One study attempts to validate a so-called "5-factor mFI"[26], but this approach violates Rockwood's own data demonstrating the need for at least 10-15 variables to have similar discrimination to the original Frailty Index[27], and with variables for only congestive heart failure, chronic obstructive pulmonary disease, hypertension, diabetes and functional status, it measures

only 2 of the 5 domains of frailty measured by the RAI. The available data regarding the RAI suggest that it can add significant value for both research and local quality improvement, and as such, consideration should be given to continued collection of standardized data through platforms like ACS NSQIP.

Another advantage of the RAI in comparison to the mFI or other frailty measures (such as the Fried or Edmonton Frail Scale) is its granular range of scores that permits users to adopt cutoffs suitable for a variety of applications depending on the prevalence of frailty in their particular population and/or the resources available for intervention. Another unique aspect of our study is that we have identified and proposed specific thresholds (Table 4 and eTables 1, 2 and 3) for identifying frailty based on the near doubling of predicted mortality in both cohorts across the spectrum of frailty risk strata. For example, with a cutoff of revised RAI-rev≥30, those identified as frail comprise the riskiest 10% of the population with a predicted mortality of at least 2.3%. A moderate resource setting such as a preoperative optimization clinic could target such a cutoff for quality improvement efforts whereas a resource limited intervention study may adopt a higher frailty threshold.

Finally, the revised RAI demonstrates extraordinarily high specificity and negative predictive values, suggesting that it could be best used as the first of a two-stage frailty screening program. In the first stage, the high NPV of the RAI is used to rapidly screen all patients and categorize them as either “robust” or “potentially frail”. Patients classified as “robust” would not require any modification to usual surgical care or decision making. The “potentially frail” would warrant a second stage in which a more critical evaluation is performed to understand and potentially rectify the precise nature of their increased risk. The precise threshold used to dichotomize patients will largely depend on both the clinical context and available resources, as

highlighted above. The revised RAI has significant precision for a wide range of scores (0-45) that encompasses a vast majority of surgical patients thus providing surgeons important data for assessment of risk and introducing a "surgical pause".

In the end, however, mere measurement of frailty-associated risks is insufficient to improve clinical care[28,29]. Clinicians must effectively communicate those risks to their patients in a process of shared decision-making that acknowledges how patient priorities can and do shift in the latter phases of life[29]. The similarity in 30-day predicted mortalities for VASQIP and ACS-NSQIP, suggests that 180-day mortality for non-veteran populations may be as high as 25-50% for RAI-rev values of 40 and above. These are important considerations for evaluating the tradeoff between survival and quality of life through elective surgery. Measurement of RAI can help patients and providers recognize when non-surgical means of palliation may be appropriate, triggering and informing deeper discussion of the goals of care. It can also help guide allocation of resources like case managers, social work, or rehabilitation to higher risk surgical patients and develop resource-effective interventions to improve quality of care[29].

This study has several important limitations. First, the scope of surgical outcomes analyzed were limited to mortality in order to effectively validate the RAI. However, the outcomes relevant for frail patients may include loss of independence, institutionalization and cognitive decline which are not captured in VASQIP or ACS NSQIP. Because these patient-centered variables are conspicuously absent from most surgical registries, this may represent an opportunity for developing and implementing new variables such as those proposed as part of the ACS Coalition for Quality in Geriatric Surgery Project[23]. Secondly, ACS NSQIP is constrained to 30-day outcomes, and as such we were unable to confirm the longer-term predictions observed in the VASQIP cohort. Finally, the role of physical performance measures like grip strength and

gait speed remains unclear, and will require further research in a prospectively enrolled cohort to tease apart the incremental improvements in predictive power afforded by these more labor-intensive assessments. Such research would also help establish the range of RAI scores that corresponds to the frailty phenotype.

In conclusion, this study provides robust recalibration and validation of the RAI in a representative sample of surgical patients from the VA and ACS-NSQIP as well as prospectively collected frailty data. The revised RAI offers improved discrimination and calibration over the original and is generalizable to US surgical populations including men and women. Based on the current analysis, the revised RAI is a precise frailty assessment tool suitable for "ruling out frailty" in a majority of surgical patients, identifying varying degrees of risk in "potentially frail" surgical patients and adaptable to various clinical settings by specifying different thresholds. As such, we propose that the RAI is sufficiently developed, calibrated and validated for implementation in surgical clinics for rapid, real-time assessment of frailty as well as for clinical use in informed consent and shared decision making with patients and providers.

REFERENCES

1. Etzioni DA, Liu JH, Maggard MA, Ko CY. The aging population and its impact on the surgery workforce. *Annals of surgery.* 2003;238(2):170-177.
2. Kwok AC, Semel ME, Lipsitz SR, et al. The intensity and variation of surgical care at the end of life: a retrospective cohort study. *Lancet.* 2011;378(9800):1408-1413.
3. Robinson TN, Wu DS, Stiegmann GV, Moss M. Frailty predicts increased hospital and six-month healthcare cost following colorectal surgery in older adults. *American journal of surgery.* 2011;202(5):511-514.
4. Schwarze ML, Brasel KJ, Mosenthal AC. Beyond 30-day mortality: Aligning surgical quality with outcomes that patients value. *JAMA surgery.* 2014;149(7):631-632.
5. Fried LP, Tangen CM, Walston J, et al. Frailty in Older Adults: Evidence for a Phenotype. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences.* 2001;56(3):M146-M157.
6. Rockwood K, Mitnitski A. Frailty defined by deficit accumulation and geriatric medicine defined by frailty. *Clinics in geriatric medicine.* 2011;27(1):17-26.
7. Sternberg SA, Wershof Schwartz A, Karunananthan S, Bergman H, Mark Clarfield A. The identification of frailty: a systematic literature review. *Journal of the American Geriatrics Society.* 2011;59(11):2129-2138.
8. Makary MA, Segev DL, Pronovost PJ, et al. Frailty as a predictor of surgical outcomes in older patients. *Journal of the American College of Surgeons.* 2010;210(6):901-908.
9. Anaya DA, Johanning J, Spector SA, et al. Summary of the panel session at the 38th Annual Surgical Symposium of the Association of VA Surgeons: what is the big deal about frailty? *JAMA Surg.* 2014;149(11):1191-1197.

10. Robinson TN, Eiseman B, Wallace JI, et al. Redefining geriatric preoperative assessment using frailty, disability and co-morbidity. *Annals of surgery.* 2009;250(3):449-455.

11. Velanovich V, Antoine H, Swartz A, Peters D, Rubinfeld I. Accumulating deficits model of frailty and postoperative mortality and morbidity: its application to a national database. *The Journal of surgical research.* 2013;183(1):104-110.

12. Gani F, Canner JK, Pawlik TM. Use of the Modified Frailty Index in the American College of Surgeons National Surgical Improvement Program Database: Highlighting the Problem of Missing Data. *JAMA Surg.* 2017;152(2):205-207.

13. Hall DE, Arya S, Schmid KK, et al. Development and Initial Validation of the Risk Analysis Index for Measuring Frailty in Surgical Populations. *JAMA Surg.* 2017;152(2):175-182.

14. Hall DE, Arya S, Schmid KK, et al. Association of a frailty screening initiative with postoperative survival at 30, 180, and 365 days. *JAMA Surg.* 2017;152(3):233-240.

15. Porock D, Parker-Oliver D, Petroski GF, Rantz M. The MDS Mortality Risk Index--Revised (MMRI-R). 2010; www.biomedcentral.com/content/supplementary/1756-0500-3-200-S1.doc.

16. American College of Surgeons National Surgical Quality Improvement Program http://site.acsnsqip.org, accessed 11/04/18. http://site.acsnsqip.org.

17. Khuri SF, Daley J, Henderson W, et al. The Department of Veterans Affairs' NSQIP: the first national, validated, outcome-based, risk-adjusted, and peer-controlled program for the measurement and enhancement of the quality of surgical care. National VA Surgical Quality Improvement Program. *Annals of surgery.* 1998;228(4):491-507.

18. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. *Annals of surgery.* 2008;248(2):329-336.

19. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 1974 Dec 19(6):716–723.

20. Nagelkerke NJD. A Note on a General Definition of the Coefficient of Determination. *Biometrika.* 1991;78(3):691–692.

21. van Walraven C, McAlister FA, Bakal JA, Hawken S, Donze J. External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne.* 2015;187(10):725-733.

22. Suskind AM, Finlayson E. A Call for Frailty Screening in the Preoperative Setting. *JAMA surgery.* 2017;152(3):240-241.

23. Berian JR, Zhou L, Hornor MA, et al. Optimizing Surgical Quality Datasets to Care for Older Adults: Lessons from the American College of Surgeons NSQIP Geriatric Surgery Pilot. *Journal of the American College of Surgeons.* 2017;225(6):702-712.e701.

24. Aguayo GA, Donneau AF, Vaillant MT, et al. Agreement Between 35 Published Frailty Scores in the General Population. *American journal of epidemiology.* 2017;186(4):420-434.

25. Adams P, Ghanem T, Stachler R, Hall F, Velanovich V, Rubinfeld I. Frailty as a predictor of morbidity and mortality in inpatient head and neck surgery. *JAMA otolaryngology-- head & neck surgery.* 2013;139(8):783-789.

26. Subramaniam S, Aalberg JJ, Soriano RP, Divino CM. New 5-Factor Modified Frailty Index Using American College of Surgeons NSQIP Data. *Journal of the American College of Surgeons.* 2018;226(2):173-181.e178.

27. Rockwood K, Andrew M, Mitnitski A. A Comparison of Two Approaches to Measuring Frailty in Elderly People. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences.* 2007;62(7):738-743.

28. Wick EC, Finlayson E. Frailty-Going From Measurement to Action. *JAMA Surg.* 2017;152(8):757-758.

29. Kopecky KE, Urbach D, Schwarze ML. Risk Calculators and Decision Aids Are Not Enough for Shared Decision Making. *JAMA Surg.* 2018.

FIGURES LEGENDS

Figure 1: Model Calibration: Observed vs. Predicted Mortality Across the Range of RAI Scores. Predicted mortality for patients undergoing elective surgery was calculated using logistic regression with RAI scores as the sole independent variable. The predicted mortality for each RAI score is plotted against the observed mortality with 95% confidence intervals. The revised RAI showed significant improvement in model calibration, as demonstrated by improved c-statistic and overlap of the predicted mortality with observed mortality.

Figure 1a: Observed vs. Predicted 180-day Mortality for RAI-A Original Score in Veterans Affairs Surgical Quality Improvement Program (VASQIP; c=0.813)

Figure 1b: Observed vs. Predicted 180-day Mortality for RAI-rev Recalibrated Score in VASQIP (c=0.842)

Figure 1c: Observed vs. Predicted 30-day Mortality for RAI-rev Recalibrated Score in American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP; c=0.87)

Figure 1d: Observed vs. Predicted 30-day Mortality for RAI-C-rev Recalibrated Score in prospectively collected data at Nebraska Western Iowa Health Care System (NWIHCS; c=0.8)

Figure 2: Model Calibration: Observed vs. Predicted 30 day Mortality Across the Range of RAI-rev Scores in (a) women [c=0.89] and (b) men [c=0.85]. Predicted mortality for patients undergoing elective surgery in the American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database was calculated using logistic regression with

RAI-rev scores as the sole independent variable. The predicted mortality for each revised RAI score is plotted against the observed mortality with 95% confidence intervals.

TABLES

Table 1. Demographic characteristics, Risk Analysis Index (RAI) components and mortality (30-, 180- and 365-day) for the recalibration cohort (development and confirmation samples) using VASQIP) data [2010-2014]; and the external validation cohort using ACS-NSQIP data [2005-2014] and NWIHCS VA data [prospective RAI-C validation].

		VASQIP		ACS-NSQIP	NWIHCS
		Recalibration Development Sample	Recalibration Confirmation Sample	External validation Sample	RAI-C validation sample
		N=322,152	N=158,579	N=1,391,785	N=6856
Variable	Category	N (%)	N (%)	N (%)	N (%)
Gender	Female	25,159 (7.8)	12,420 (7.8)	807,087 (58.0)	249 (3.6)
	Male	296,993 (92.2)	146,159 (92.2)	584,698 (42.0)	6,607 (96.4)
Age at Time of RAI (yrs.)	< 20	101 (0.0)	56 (0.0)	17,369 (1.3)	1 (0.0)
	20-24	2,086 (0.7)	1,067 (0.7)	28,802 (2.1)	64 (0.9)
	25-29	7,012 (2.2)	3,396 (2.1)	47,097 (3.4)	193 (2.8)
	30-34	7,858 (2.4)	3,811 (2.4)	61,612 (4.4)	197 (2.9)
	35-39	7,737 (2.4)	3,760 (2.4)	80,860 (5.8)	186 (2.7)
	40-44	12,285 (3.8)	6,015 (3.8)	105,259 (7.6)	256 (3.7)
	45-49	17,305 (5.4)	8,613 (5.4)	131,351 (9.4)	354 (5.2)
	50-54	28,834 (9.0)	14,316 (9.0)	149,172 (10.7)	570 (8.3)
	55-59	40,738 (12.7)	20,288 (12.8)	153,204 (11.0)	712 (10.4)
	60-64	75,598 (23.5)	36,971 (23.3)	153,597 (11.0)	1,458 (21.3)
	65-69	55,852 (17.3)	27,163 (17.1)	142,281 (10.2)	1,299 (19.0)
	70-74	24,845 (7.7)	12,377 (7.8)	115,397 (8.3)	601 (8.8)
	75-79	19,738 (6.1)	9,749 (6.2)	94,480 (6.8)	443 (6.5)
	80-84	13,153 (4.1)	6,474 (4.1)	66,581 (4.8)	332 (4.8)
	85-89	7,040 (2.2)	3,530 (2.2)	33,460 (2.4)	157 (2.3)
	≥ 90	1,970 (0.6)	993 (0.6)	11,263 (0.8)	33 (0.5)
	Mean (SD)	*60.7 (13.1)*	*60.7 (13.1)*	*56.3 (16.6)*	*60.7 (13.9)*
Race	White	222,722 (69.1)	109,535 (69.1)	912,117 (75.6)	2,224 (32.4)
	Black	48,906 (15.2)	24,154 (15.2)	119,002 (9.9)	128 (1.9)
	Other	3,586 (1.1)	1,725 (1.1)	35,503 (2.9)	4,189 (61.1)
	Unknown	46,938 (14.6)	23,165 (14.6)	140,381 (11.6)	315 (4.6)
Ethnicity	Not Hispanic or Latino	281,930 (87.5)	139,125 (87.7)	1,060,916 (76.2)	6,509 (94.9)
	Hispanic or Latino	17,119 (5.3)	8,187 (5.2)	77,516 (5.6)	32 (0.5)
	Unknown	23,023 (7.2)	11,267 (7.1)	253,353 (18.2)	315 (4.6)

BMI[1]	**< 18.5**	5,906 (1.8)	3,015 (1.9)	20720 (1.5)	6 (0.1)
	≥ 18.5 & < 25	78,434 (24.4)	38,182 (24.1)	317,026 (22.8)	103 (1.5)
	≥ 25 & < 30	113,635 (35.3)	55,976 (35.3)	420,945 (30.2)	200 (2.9)
	≥ 30 & < 35	77,621 (24.1)	38,380 (24.2)	281,999 (20.3)	152 (2.2)
	≥ 35 & < 40	31,542 (9.8)	15,560 (9.8)	150,839 (10.8)	91 (1.3)
	≥ 40	13,675 (4.2)	6,758 (4.3)	177,213 (12.7)	28 (0.6)
	Unknown	1,339 (0.4)	709 (0.4)	23,043 (1.7)	6,266 (91.4)
	Mean (SD)	*28.9 (5.9)*	*28.9 (5.9)*	*30.8 (8.4)*	*30.3 (6.3)*
Cancer[#]		6,341 (2.0)	3,175 (2.0)	26,773 (1.9)	1,197 (17.5)
Unintentional weight loss		8,180 (2.5)	4,057 (2.6)	26,391 (1.9)	273 (4.0)
Poor Appetite		NA	NA	NA	312 (4.6)
Renal Failure		1,461 (0.5)	777 (0.5)	26,412 (1.9)	239 (3.5)
Congestive Heart Failure		2,189 (0.7)	1,103 (0.7)	10,297 (0.7)	274 (4.0)
Dyspnea		3,421 (1.1)	1,682 (1.1)	12,610 (0.9)	433 (3.3)
Other Living Setting		4,724 (1.5)	2,314 (1.5)	32,713 (2.4)	157 (2.3)
Cognitive Decline		11,343 (3.5)	5,571 (3.5)	33,443 (2.4)	93 (1.4)
Functional Status	**Independent**	298,590 (92.7)	146,959 (92.7)	1,329,567 (95.5)	NA
	Partially Dependent	18,320 (5.7)	8,990 (5.7)	49,919 (3.6)	NA
	Totally Dependent	5,242 (1.6)	2,630 (1.7)	12,299 (0.9)	NA
Activities of Daily Living: Mobility/Locomotion	**0:Independent**	NA	NA	NA	6,692 (97.6)
	1: Supervised	NA	NA	NA	35 (0.5)
	2: Limited Assist	NA	NA	NA	69 (1.0)
	3: Extensive Assist	NA	NA	NA	32 (0.5)
	4: Total Dependent	NA	NA	NA	28 (0.4)
Activities of Daily Living: Eating	**0: Independent**	NA	NA	NA	6,778 (98.9)
	1: Supervised	NA	NA	NA	30 (0.4)
	2: Limited Assist	NA	NA	NA	18 (0.3)
	3: Extensive Assist	NA	NA	NA	11 (0.2)
	4: Total Dependent	NA	NA	NA	19 (0.3)
Activities of Daily Living: Toilet Use	**0: Independent**	NA	NA	NA	6,749 (98.4)
	1: Supervised	NA	NA	NA	26 (0.4)
	2: Limited Assist	NA	NA	NA	32 (0.5)
	3: Extensive Assist	NA	NA	NA	23 (0.3)
	4: Total Dependent	NA	NA	NA	26 (0.4)
Activities of Daily Living: Personal Hygiene	**0: Independent**	NA	NA	NA	6,724 (98.1)
	1: Supervised	NA	NA	NA	31 (0.4)
	2: Limited Assist	NA	NA	NA	50 (0.7)
	3: Extensive Assist	NA	NA	NA	25 (0.4)

	4: Total Dependent	NA	NA	NA	26 (0.4)
RAI	**0-4**	23,410 (7.3)	11,543 (7.3)	688,422 (49.5)	203 (3.0)
	5-9	244,255 (75.8)	119,779 (75.5)	523,123 (37.6)	4,280 (62.4)
	10-14	18,857 (5.9)	9,582 (6.0)	83,116 (6.0)	593 (8.7)
	15-19	17,946 (5.6)	8,871 (5.6)	44,104 (3.2)	417 (6.1)
	20-24	6,454 (2.0)	3,191 (2.0)	25,048 (1.8)	461 (6.7)
	25-29	5,779 (1.8)	2,868 (1.8)	16,509 (1.2)	631 (9.2)
	30-34	3,197 (1.0)	1,662 (1.1)	6,583 (0.5)	177 (2.6)
	≥ 35	2,254 (0.7)	1,083 (0.7)	4,880 (0.4)	94 (1.4)
	Mean (SD)	*8.6 (5.6)*	*8.7 (5.6)*	*5.9 (5.5)*	*11.8 (8.1)*
Revised RAI	**0-4**	2,753 (0.9)	1,402 (0.9)	73,594 (5.3)	80 (1.2)
	5-9	19,418 (6.0)	9,432 (6.0)	127,813 (9.2)	367 (5.4)
	10-14	24,572 (7.6)	12,108 (7.6)	289,473 (20.8)	443 (6.5)
	15-19	71,150 (22.1)	34,951 (22.0)	315,783 (22.7)	1,286 (18.8)
	20-24	122,687 (38.1)	59,829 (37.7)	293,327 (21.1)	1,993 (29.1)
	25-29	46,504 (14.4)	23,460 (14.8)	180,474 (13.0)	1,010 (14.7)
	30-34	17,224 (5.4)	8,637 (5.5)	61,669 (4.4)	354 (5.2)
	35-39	10,611 (3.3)	5,159 (3.3)	31,534 (2.3)	979 (14.3)
	40-44	4,433 (1.4)	2,203 (1.4)	11,552 (0.8)	210 (3.1)
	45-49	1,683 (0.5)	870 (0.6)	4,633 (0.3)	96 (1.4)
	≥ 50	1,117 (0.4)	528 (0.3)	1,933 (0.1)	38 (0.6)
	Mean (SD)	*21.2 (7.5)*	*21.2 (7.5)*	*18.1 (8.4)*	*23.9 (9.4)*
30-day Mortality after Surgery		3,463 (1.1)	1,801 (1.1)	13,408 (1.0)	29 (0.4)
180-day Mortality after Surgery		11,416 (3.5)	5,835 (3.7)	NA	114 (1.8)

VASQIP: Veterans Affairs Surgical Quality Improvement Program; ACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program; NWIHCS: Nebraska Western Iowa Health Care System; SD: standard deviation; BMI: body mass index

[1]Values below 10 and above 90 were recoded to unknown

Table 2. Model parameters for recalibration of the Risk Analysis Index (RAI) comparing origin RAI (RAI-A) to the recalibrated RAI (RAI-rev) in VASQIP dataset; external validation of RAI-rev in ACS-NSQIP cohort and subcohorts of men and women.

Outcome	Sample	Predictor	C-statistic (95% C.I.)	AIC	Max. R^2
VASQIP Recalibration					
30-day Mortality	Total [N=480,731]	RAI-A	0.842 (0.835-0.848)	47,002.2	0.1990
		RAI-rev	0.864 (0.858-0.869)	45,104.2	0.2330
180-day Mortality	Total [N=480,731]	RAI-A	0.813 (0.810-0.817)	120,967.0	0.2110
		RAI-rev	0.842 (0.839-0.845)	114,881.8	0.2550
365-day Mortality	Total [N=480,731]	RAI-A	0.784 (0.781-0.787)	175,931.5	0.1970
		RAI-rev	0.816 (0.814-0.819)	167,259.0	0.2440
ACS-NSQIP External Validation					
30-day Mortality	Female [N=807,087]	RAI-rev	0.885 (0.881-0.889)	55,462.3	0.2398
	Male [N=584,698]	RAI-rev	0.845 (0.841- 0.850)	63,368.8	0.1999
	Total [N=1,391,785]	RAI-rev	0.870 (0.867-0.873)	118,997.0	0.2221
RAI-C NWIHCS Validation					
30-day Mortality	Total [N=6,803]	RAI-C	0.704 (0.596-0.812)	356.8	0.0590
		RAI-C- rev	0.743 (0.657-0.829)	353.4	0.0690
180-day Mortality	Total [N=6,419]	RAI-C	0.772 (0.727-0.816)	1,030.1	0.1120
		RAI-C-rev	0.804 (0.766-0.842)	1,000.7	0.1400

p<0.0001 for all model comparisons between original RAI and RAI-rev in VASQIP and ACS-NSQIP cohorts; p=0.204 for 30-day mortality and p<.001 for 180-day mortality in the NWIHCS sample;
VASQIP: Veterans Affairs Surgical Quality Improvement Program; ACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program; NWIHCS: Nebraska Western Iowa Health Care System; CI: Confidence interval; AIC: Aikake information criterion

Table 3. Revised Risk Analysis Index scoring for the prospective (RAI-C) and retrospective (RAI-A) versions.

Variable	Revised RAI-A		Revised RAI-C	
Sex	3		3	
Age*Cancer	w/o cancer	w/ cancer	w/o cancer	w/ cancer
Age				
<=19	0	28	0	28
20-24	1	29	1	29
25-29	4	29	4	29
30-34	6	30	6	30
35-39	8	30	8	30
40-44	10	31	10	31
45-49	12	31	12	31
50-54	14	32	14	32
55-59	16	32	16	32
60-64	18	33	18	33
65-69	20	34	20	34
70-74	22	34	22	34
75-79	24	35	24	35
80-84	26	35	26	35
85-89	28	36	28	36
90-94	30	36	30	36
95-99	32	37	32	37
100+	34	37	34	37
Weight Loss	4		4	
Poor Appetite	4		4	
Renal Failure	8		8	
Chronic/Congestive Heart Failure	5		5	
Shortness of Breath	3		3	
Residence other than Ind. Living	1		1	
ADL*Cog	w/o cog	w/cog	w/o cog	w/ cog
Totally dependent	14	16		
Partially dependent	7	11		
Independent	0	5		
ADL Score				
0			0	5
1			1	6
2			2	6
3			3	7
4			4	8
5			4	8
6			5	9
7			6	10
8			7	11

9			8	11
10			9	12
11			10	13
12			11	13
13			11	14
14			12	15
15			13	15
16			14	16
Total RAI	0	81	0	81

Table 4: Proposed thresholds for clinical use for the recalibrated Risk analysis Index (RAI-rev). Frailty prevalence, negative and positive predictive values, sensitivity and specificity, predicted 30-day mortality and predicted 180-day mortality presented for VASQIP and ACS-NSQIP cohorts for each proposed RAI-rev thresholds of 25, 30, 35, 40 and 45. Similar statistics for the NWIHCS cohort for proposed RAI-C-rev thresholds of 30, 37, 45 and 53.

Revised RAI threshold	Frailty Prevalence	Negative Predictive Value	Positive Predictive Value	Sensitivity	Specificity	Observed 30- day Mortality Rate	Predicted 30- day Mortality Rate	Predicted 180- day Mortality Rate
VASQIP								
25	25.5%	98.9%	10.8%	76.9%	76.5%	0.9%	0.8%	3.1%
30	10.9%	98.4%	19.4%	59.0%	90.9%	2.3%	1.8%	6.7%
35	5.5%	97.8%	27.9%	43.1%	95.9%	4.2%	3.8%	13.9%
40	2.3%	97.3%	40.6%	25.5%	98.6%	12.2%	7.9%	26.6%
45	0.9%	96.8%	51.6%	12.6%	99.6%	16.7%	15.9%	44.8%
ACS-NSQIP								
25	20.0%	99.7%	3.6%	77.8%	79.6%	1.0%	1.1%	NA
30	10.0%	99.6%	7.0%	58.0%	92.5%	3.6%	2.3%	NA
35	5.8%	99.4%	11.0%	40.9%	96.8%	5.4%	4.9%	NA
40	3.9%	99.3%	17.9%	24.2%	98.9%	12.9%	10.2%	NA
45	3.2%	99.1%	24.6%	12.1%	99.6%	17.8%	20.1%	NA
NWIHCS (RAI-C-rev)								
30	23.6%	99.2%	4.9%	64.9%	77.2%	0.0%	0.5%	2.0%
37	14.0%	99.0%	6.4%	50.0%	86.7%	0.3%	0.9%	4.3%
45	1.9%	98.5%	14.6%	15.8%	98.3%	2.9%	2.0%	10.3%
53	0.3%	98.3%	23.5%	3.5%	99.8%	12.5%	4.0%	22.4%

VASQIP: Veterans Affairs Surgical Quality Improvement Program; ACS-NSQIP: American College of Surgeons National Surgical Quality Improvement Program; NWIHCS: Nebraska Western Iowa Health Care System;

Figure 1

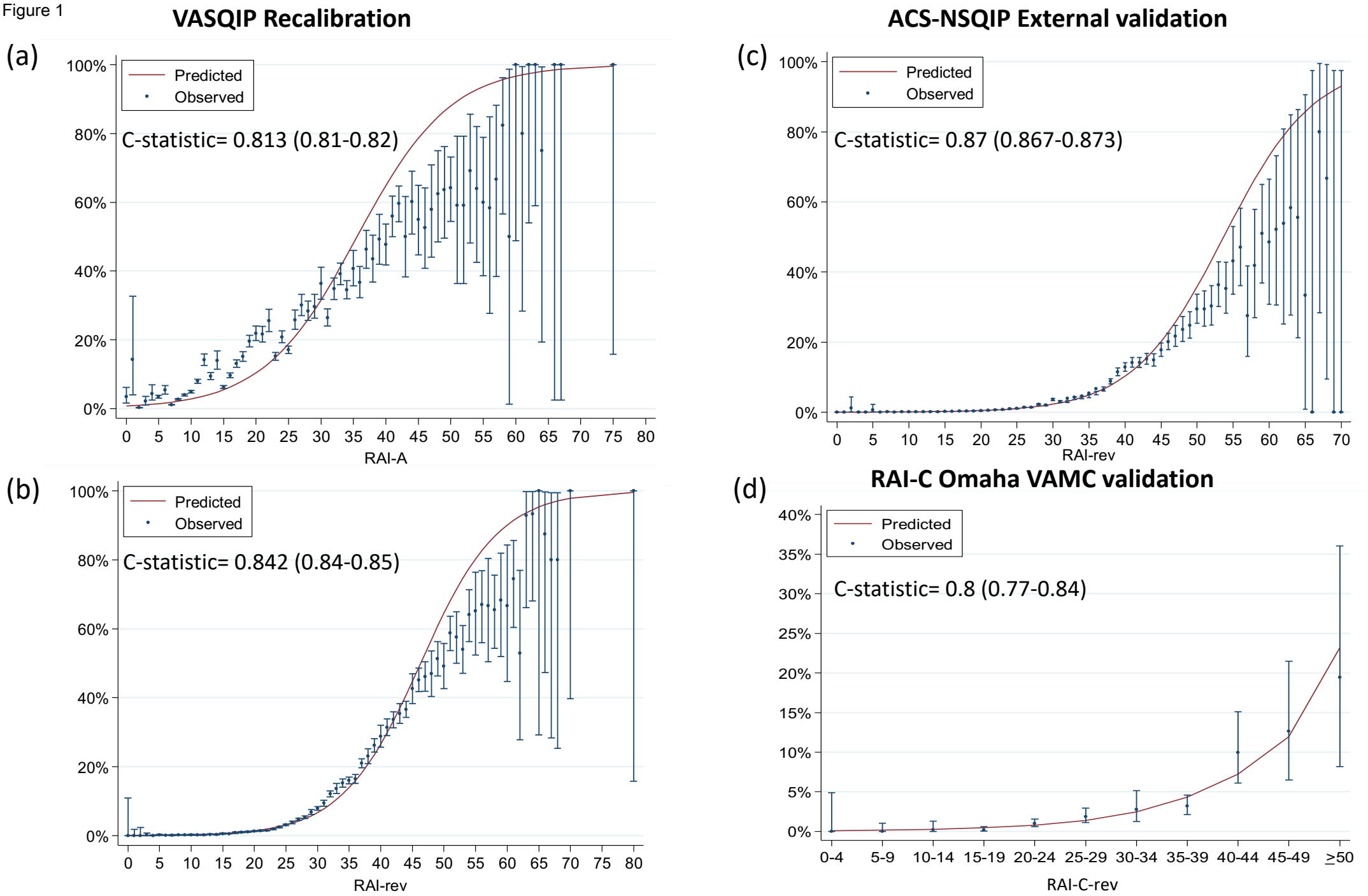
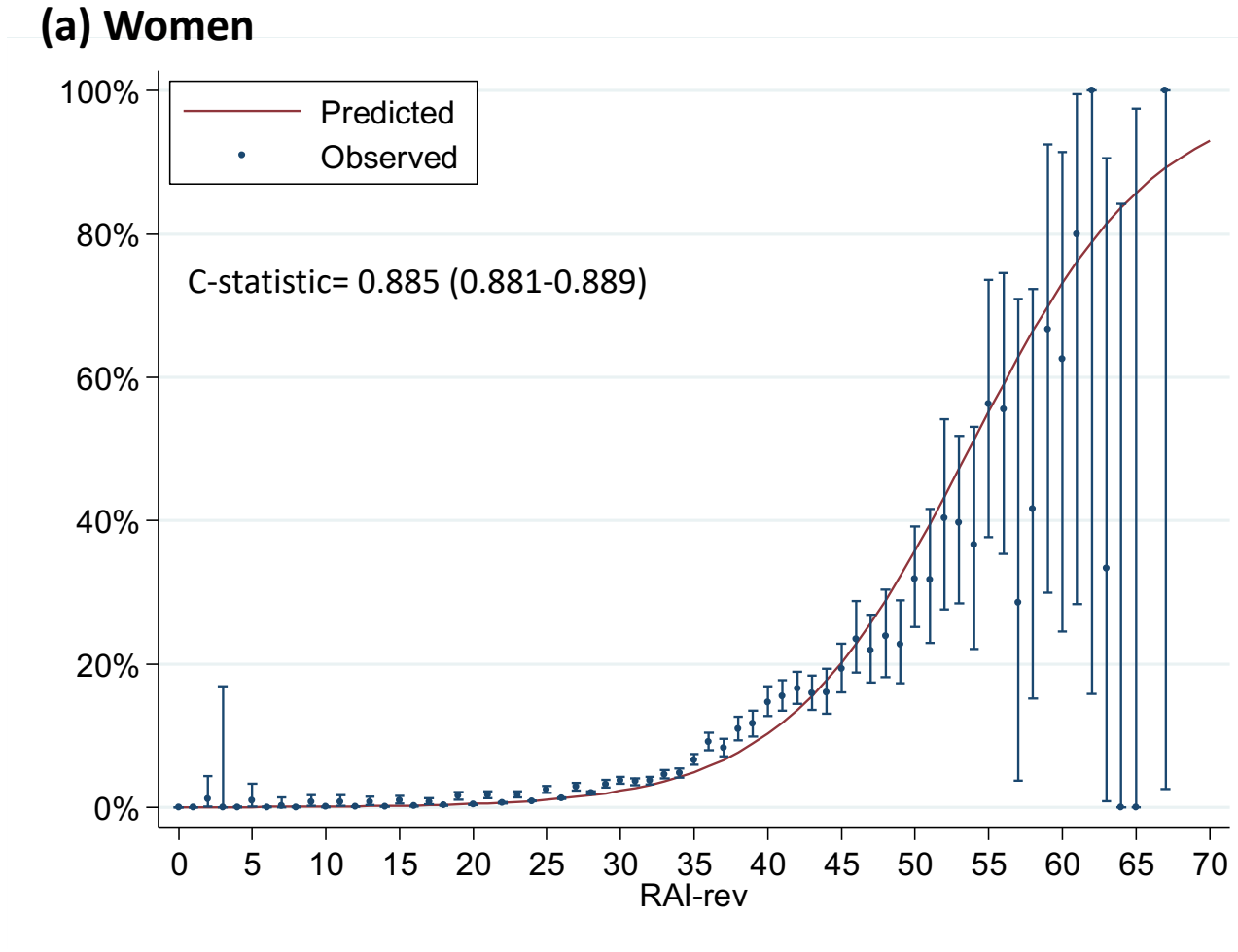
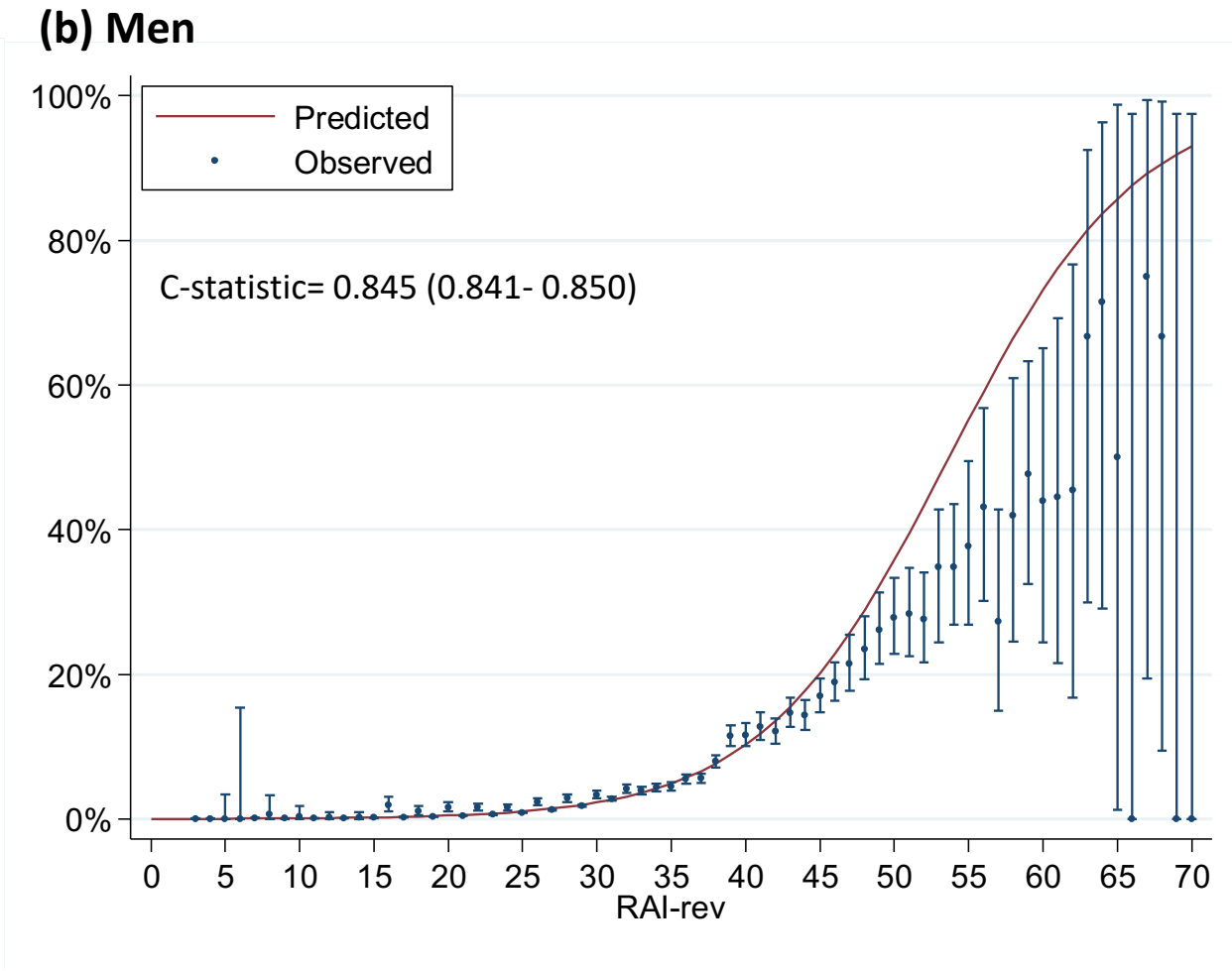

VASQIP Recalibration
(a)
Predicted
Observed
C-statistic= 0.813 (0.81-0.82)
100%
80%
60%
40%
20%
0%
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
RAI-A
(b)
Predicted
Observed
C-statistic= 0.842 (0.84-0.85)
100%
80%
60%
40%
20%
0%
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80
RAI-rev
ACS-NSQIP External validation
(c)
Predicted
Observed
C-statistic= 0.87 (0.867-0.873)
100%
80%
60%
40%
20%
0%
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
RAI-rev
RAI-C Omaha VAMC validation
(d)
Predicted
Observed
C-statistic= 0.8 (0.77-0.84)
40%
35%
30%
25%
20%
15%
10%
5%
0%
0-4 5-9 10-14 15-19 20-24 25-29 30-34 35-39 40-44 45-49 ≥50
RAI-C-rev

Figure 2


(a) Women
Predicted
Observed
C-statistic= 0.885 (0.881-0.889)
100%
80%
60%
40%
20%
0%
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
RAI-rev


(b) Men
Predicted
Observed
C-statistic= 0.845 (0.841- 0.850)
100%
80%
60%
40%
20%
0%
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70
RAI-rev

Appendix

eTable 1: Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, observed and predicted 30-day and 180-day mortality for revised Risk Analysis Index (RAI) in Veterans Affairs Surgical Quality Improvement Program (VASQIP) datasets (2010-2014)

RAI-rev Threshold	N	Frailty Prevalence (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	Observed 30-day mortality (%)	Predicted 30-Day Mortality (%)	Observed 180-day mortality (%)	Predicted 180-Day Mortality (%)
0	32	100	100	0	3.59		0	0.02	0	0.06
1	196	99.99	100	0.01	3.59	100	0	0.02	0	0.07
2	151	99.95	100	0.05	3.59	100	0	0.02	0	0.08
3	1028	99.92	100	0.08	3.59	100	0	0.03	0.19	0.09
4	2748	99.71	99.99	0.3	3.6	99.86	0	0.03	0.04	0.11
5	3453	99.14	99.98	0.9	3.62	99.93	0.06	0.04	0.2	0.13
6	6708	98.42	99.94	1.64	3.64	99.87	0	0.04	0.12	0.15
7	2988	97.02	99.9	3.08	3.69	99.87	0.03	0.05	0.13	0.18
8	8716	96.4	99.87	3.73	3.72	99.87	0.03	0.06	0.21	0.21
9	6985	94.59	99.77	5.61	3.79	99.85	0	0.07	0.16	0.24
10	2440	93.13	99.7	7.11	3.84	99.85	0.04	0.08	0.2	0.29
11	8428	92.63	99.68	7.64	3.86	99.84	0.06	0.09	0.17	0.34
12	7302	90.87	99.59	9.45	3.93	99.84	0.05	0.11	0.16	0.39
13	12758	89.35	99.52	11.02	4	99.84	0.07	0.13	0.37	0.46
14	5752	86.7	99.25	13.77	4.11	99.8	0.05	0.15	0.3	0.54
15	16432	85.5	99.15	15	4.16	99.79	0.18	0.17	0.59	0.64
16	28794	82.09	98.59	18.53	4.31	99.72	0.1	0.2	0.59	0.75
17	8893	76.1	97.61	24.7	4.6	99.64	0.31	0.23	0.85	0.88
18	29710	74.25	97.17	26.61	4.7	99.61	0.22	0.27	1.02	1.03
19	22272	68.07	95.41	32.95	5.03	99.48	0.28	0.32	1.1	1.21
20	48186	63.43	93.99	37.7	5.32	99.41	0.35	0.37	1.3	1.42
21	23107	53.41	90.36	47.97	6.07	99.26	0.39	0.44	1.44	1.66
22	60950	48.6	88.43	52.88	6.53	99.19	0.36	0.51	1.55	1.95
23	35004	35.92	82.97	65.83	8.29	99.05	0.44	0.6	1.91	2.28

24	15269	28.64	79.09	73.23	9.91	98.95	0.69	0.7	2.49	2.67
25	19679	25.47	76.89	76.45	10.83	98.89	0.85	0.82	3.11	3.12
26	11359	21.37	73.33	80.56	12.31	98.78	0.99	0.95	3.81	3.64
27	18517	19.01	70.82	82.92	13.37	98.71	1.25	1.11	4.76	4.26
28	15369	15.16	65.71	86.72	15.56	98.55	1.19	1.3	5.33	4.96
29	5040	11.96	60.96	89.86	23.54	98.41	1.69	1.52	6.87	5.78
30	9912	10.91	58.96	90.87	19.39	98.35	2.26	1.77	7.92	6.73
31	4446	8.85	54.41	92.84	22.06	98.21	2.32	2.06	9.38	7.81
32	5547	7.93	51.99	93.71	23.54	98.13	3.28	2.4	12.08	9.05
33	2181	6.77	48.11	94.77	25.49	98	3.9	2.8	13.62	10.47
34	3775	6.32	46.39	95.17	26.34	97.95	4.24	3.26	15.23	12.09
35	5455	5.53	43.05	95.86	27.92	97.84	4.23	3.79	15.97	13.91
36	3051	4.4	38	96.85	31	97.67	3.97	4.41	16.42	15.95
37	3739	3.76	35.1	97.4	33.46	97.58	6.45	5.12	20.97	18.24
38	1499	2.99	30.55	98.04	36.71	97.43	8.61	5.93	23.02	20.77
39	2026	2.68	28.55	98.29	38.3	97.37	9.62	6.87	26.16	23.54
40	820	2.25	25.48	98.61	40.58	97.26	12.2	7.94	28.78	26.57
41	1420	2.08	24.11	98.74	41.54	97.22	11.97	9.17	31.41	29.83
42	1618	1.79	21.53	98.95	43.22	97.13	10.38	10.56	33.62	33.32
43	1070	1.45	18.38	99.18	45.44	97.03	10.93	12.13	35.42	36.99
44	1708	1.23	16.18	99.33	47.26	96.95	13.64	13.9	36.59	40.82
45	528	0.87	12.56	99.56	51.6	96.83	16.67	15.88	42.61	44.77
46	848	0.76	11.25	99.63	52.89	96.79	21.93	18.09	45.17	48.78
47	542	0.59	9.03	99.73	55.21	96.72	22.32	20.53	46.13	52.81
48	230	0.47	7.58	99.79	57.37	96.67	26.96	23.2	46.96	56.81
49	405	0.43	6.96	99.82	58.54	96.65	24.44	26.11	51.36	60.71
50	238	0.34	5.75	99.86	60.3	96.61	18.49	29.25	49.16	64.49
51	388	0.29	5.07	99.89	62.19	96.58	28.61	32.59	58.76	68.09
52	177	0.21	3.75	99.92	63.49	96.54	28.25	36.12	57.63	71.49
53	209	0.18	3.16	99.94	64.73	96.52	25.84	39.81	54.07	74.66
54	167	0.13	2.5	99.96	68.25	96.5	30.54	43.62	64.07	77.59
55	66	0.1	1.88	99.97	69.74	96.48	36.36	47.5	65.15	80.27
56	85	0.08	1.63	99.97	70.5	96.47	43.53	51.41	67.06	82.7

57	42	0.07	1.3	99.98	71.43	96.46	33.33	55.31	66.67	84.89
58	84	0.06	1.14	99.98	72.16	96.45	27.38	59.14	65.48	86.84
59	41	0.04	0.82	99.99	75.13	96.44	51.22	62.87	68.29	88.58
60	24	0.03	0.66	99.99	77.03	96.43	41.67	66.45	66.67	90.11
61	51	0.03	0.57	99.99	79.03	96.43	39.22	69.85	74.51	91.46
62	17	0.02	0.35	100	82.19	96.42	41.18	73.04	52.94	92.64
63	14	0.01	0.3	100	91.07	96.42	42.86	76.01	92.86	93.67
64	15	0.01	0.22	100	90.48	96.42	60	78.75	93.33	94.56
65	3	0.01	0.14	100	88.89	96.42	0	81.25	100	95.33
66	8	0.00005	0.12	100	87.5	96.42	87.5	83.52	87.5	96
67	5	0.00003	0.08	100	87.5	96.41	60	85.57	80	96.57
68	5	0.00002	0.06	100	90.91	96.41	40	87.4	80	97.07
70	4	0.00001	0.03	100	100	96.41	50	90.46	100	97.86
80	2	0.000004	0.01	100	100	96.41	0	87.85	100	99.57

eTable 2: Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, observed and predicted 30-day mortality for revised Risk Analysis Index (RAI) in American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) datasets (2005-2014)

RAI-rev Threshold	N	Frailty Prevalence (%)	Sensitivity (%)	Specificity (%)	Positive Predictive Value (%)	Negative Predictive Value (%)	Observed 30-day mortality (%)	Predicted 30-Day Mortality (%)
0	9,484	100	100	0	1	-	0.0	0
1	16,485	99.3	100	0.7	1	100	0.0	0
2	165	98.1	100	1.9	1	100	1.2	0
3	7,091	98.1	100	1.9	1	100	0.0	0
4	40,009	97.6	100	2.4	1	100	0.0	0
5	325	94.7	99.9	5.3	1	100	0.6	0
6	39,645	94.7	99.9	5.4	1	100	0.0	0.1
7	15,765	93.6	99.8	8.2	1	100	0.1	0.1
8	52,414	89.8	99.6	9.4	1.1	100	0.0	0.1
9	19,664	88.4	99.5	13.2	1.1	100	0.1	0.1
10	66,272	83.6	99.4	14.6	1.1	100	0.1	0.1
11	25,423	81.8	99.1	19.4	1.2	100	0.1	0.1
12	79,957	76	98.9	21.3	1.2	99.9	0.1	0.1
13	33,856	73.6	98.4	27	1.3	99.9	0.1	0.2
14	83,965	67.6	98.1	29.5	1.3	99.9	0.1	0.2
15	44,256	64.4	97.4	35.6	1.4	99.9	0.2	0.2
16	79,427	58.7	96.8	38.8	1.5	99.9	0.2	0.3
17	54,733	54.8	95.6	44.5	1.6	99.9	0.3	0.3
18	76,211	49.3	94.5	48.5	1.8	99.9	0.3	0.3
19	61,156	44.9	92.9	54	1.9	99.9	0.4	0.4
20	67,425	40	91.3	58.4	2.1	99.9	0.4	0.5
21	64,117	35.4	89.3	63.3	2.3	99.8	0.5	0.6
22	54,202	31.5	86.8	67.9	2.6	99.8	0.7	0.7
23	62,694	27	84.2	71.8	2.8	99.8	0.7	0.8
24	44,889	23.8	80.9	76.4	3.2	99.8	0.9	0.9

25	52,819	20	77.8	79.6	3.6	99.7	1.0	1.1
26	32,829	17.7	73.9	83.4	4.1	99.7	1.4	1.2
27	43,339	14.5	70.5	85.7	4.6	99.7	1.4	1.4
28	19,676	13.1	66	88.8	5.4	99.6	2.2	1.7
29	31,811	10.8	62.7	90.2	5.9	99.6	2.0	2
30	11,466	10	58	92.5	7	99.6	3.6	2.3
31	19,236	8.6	54.9	93.3	7.4	99.5	3.0	2.7
32	9,994	7.9	50.6	94.6	8.4	99.5	3.9	3.1
33	11,072	7.1	47.7	95.3	9.1	99.5	4.2	3.6
34	9,901	6.4	44.2	96.1	10	99.4	4.5	4.2
35	9,195	5.8	40.9	96.8	11	99.4	5.4	4.9
36	6,879	5.3	37.1	97.4	12.3	99.4	6.6	5.7
37	7,137	4.7	33.7	97.9	13.5	99.3	6.4	6.6
38	5,154	4.4	30.3	98.4	15.4	99.3	8.8	7.7
39	3,169	4.2	27	98.7	17	99.3	11.5	8.9
40	2,857	3.9	24.2	98.9	17.9	99.3	12.9	10.2
41	2,430	3.8	21.5	99.1	18.9	99.2	14.2	11.8
42	2,443	3.6	18.9	99.3	19.8	99.2	14.1	13.6
43	2,113	3.4	16.3	99.4	21.1	99.2	15.2	15.5
44	1,709	3.3	14	99.5	22.6	99.2	14.9	17.7
45	1,598	3.2	12.1	99.6	24.6	99.1	17.8	20.1
46	1,145	3.1	9.9	99.7	26.8	99.1	20.1	22.8
47	768	3.1	8.2	99.8	28.9	99.1	21.6	25.7
48	585	3	7	99.8	30.7	99.1	23.6	28.9
49	537	3	6	99.9	32.3	99.1	24.8	32.2
50	480	3	5	99.9	34.5	99.1	29.4	35.8
51	330	2.9	3.9	99.9	36.1	99.1	29.4	39.5
52	271	2.9	3.2	99.9	38.1	99.1	30.3	43.4
53	231	2.9	2.6	100	40.6	99.1	36.4	47.3
54	176	2.9	2	100	42.2	99.1	35.2	51.2
55	109	2.9	1.5	100	44.9	99.1	43.1	55.2
56	85	2.9	1.1	100	45.5	99	47.1	59.1
57	51	2.9	0.8	100	45	99	27.5	62.8
58	43	2.9	0.7	100	49.5	99	41.9	66.5

59	53	2.9	0.6	100	51.6	99	50.9	69.9
60	33	2.9	0.4	100	51.9	99	48.5	73.1
61	23	2.9	0.3	100	53.5	99	52.2	76.1
62	13	2.9	0.2	100	54.2	99	53.9	78.9
63	12	2.9	0.1	100	54.3	99	58.3	81.4
64	9	2.8	0.1	100	52.2	99	55.6	83.7
65	3	2.8	0.1	100	50	99	33.3	85.7
66	1	2.8	0	100	54.5	99	0.0	87.6
67	5	2.8	0	100	60	99	80.0	89.2
68	3	2.8	0	100	40	99	66.7	90.6
69	1	2.8	0	100	0	99	0.0	91.9
70	1	2.8	0	100	0	99	0.0	93

eTable 3: Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, observed and predicted 180-day mortality for revised Risk Analysis Index (RAI-C-rev) in Nebraska Western Iowa Health Care System (NWIHCS) prospectively collected dataset (2011-2015)

RAI-C-r Threshold	Number of Patients (% within Total)	Frailty Prevalence (%)	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Observed 30-day mortality (% within RAI-C-r)	Predicted 30-Day Mortality (Logit)	Number of 180-Day Mortalities (% within RAI-C-r)	Predicted 180-Day Mortality Rate (Logit)
1	7 (0.1)	6,419 (100)	100	0	1.8	0	0 (0.0)	0.0	0 (0.0)	0.1
3	1 (0.0)	6,412 (99.9)	100	0.1	1.8	100	0 (0.0)	0.0	0 (0.0)	0.1
4	66 (1.0)	6,411 (99.9)	100	0.1	1.8	100	0 (0.0)	0.0	0 (0.0)	0.1
6	26 (0.4)	6,345 (99.9)	100	1.2	1.8	100	0 (0.0)	0.1	0 (0.0)	0.1
7	153 (2.4)	6,319 (98.4)	100	1.6	1.8	100	0 (0.0)	0.1	0 (0.0)	0.1
8	20 (0.3)	6,166 (96.1)	100	4	1.9	100	0 (0.0)	0.1	0 (0.0)	0.2
9	151 (2.4)	6,146 (95.8)	100	4.3	1.9	100	0 (0.0)	0.1	0 (0.0)	0.2
10	23 (0.4)	5,995 (93.4)	100	6.7	1.9	100	0 (0.0)	0.1	0 (0.0)	0.2
11	143 (2.2)	5,972 (93.0)	100	7.1	1.9	100	0 (0.0)	0.1	0 (0.0)	0.2
12	26 (0.4)	5,829 (90.8)	100	9.4	2	100	1 (3.7)	0.1	1 (3.9)	0.3
13	204 (3.2)	5,803 (90.4)	99.1	9.7	2	99.8	0 (0.0)	0.1	0 (0.0)	0.3
14	28 (0.4)	5,599 (87.2)	99.1	13	2	99.9	0 (0.0)	0.1	0 (0.0)	0.3
15	260 (4.1)	5,571 (86.8)	99.1	13.4	2	99.9	0 (0.0)	0.1	0 (0.0)	0.4
16	28 (0.4)	5,311 (82.7)	99.1	17.6	2.1	99.9	0 (0.0)	0.1	0 (0.0)	0.4
17	430 (6.7)	5,283 (82.3)	99.1	18	2.1	99.9	0 (0.0)	0.2	1 (0.2)	0.5
18	34 (0.5)	4,853 (75.6)	98.3	24.8	2.3	99.9	0 (0.0)	0.2	0 (0.0)	0.5
19	464 (7.2)	4,819 (75.1)	98.3	25.3	2.3	99.9	0 (0.0)	0.2	1 (0.2)	0.6
20	31 (0.5)	4,355 (67.9)	97.4	32.7	2.6	99.9	1 (2.9)	0.2	0 (0.0)	0.6
21	947 (14.8)	4,324 (67.4)	97.4	33.2	2.6	99.9	1 (0.1)	0.2	8(0.8)	0.7
22	47 (0.7)	3,377 (52.6)	90.4	48.1	3.1	99.6	0 (0.0)	0.2	1 (2.1)	0.8
23	776 (12.1)	3,330 (51.9)	89.5	48.8	3.1	99.6	5 (0.6)	0.3	8 (1.0)	0.9
24	83 (1.3)	2,554 (39.8)	82.5	61	3.7	99.5	2 (2.3)	0.3	2 (2.4)	1
25	359 (5.6)	2,471 (38.5)	80.7	62.3	3.7	99.4	0 (0.0)	0.3	3 (0.8)	1.1
26	74 (1.2)	2,112 (32.9)	78.1	67.9	4.2	99.4	1 (1.3)	0.3	3 (4.1)	1.3
27	272 (4.2)	2,038 (31.8)	75.4	69	4.2	99.4	0 (0.0)	0.4	3 (1.1)	1.4

28	62 (1.0)	1,766 (27.5)	72.8	73.3	4.7	99.3	0 (0.0)	0.4	1 (1.6)	1.6
29	190 (3.0)	1,704 (26.6)	71.9	74.3	4.8	99.3	2 (1.0)	0.5	8 (4.2)	1.8
30	57 (0.9)	1,514 (23.6)	64.9	77.2	4.9	99.2	0 (0.0)	0.5	1 (1.8)	2
31	110 (1.7)	1,457 (22.7)	64	78.1	5	99.2	0 (0.0)	0.5	1 (0.9)	2.2
32	49 (0.8)	1,347 (21.0)	63.2	79.8	5.4	99.2	1 (1.9)	0.6	3 (6.1)	2.5
33	56 (0.9)	1,298 (20.2)	60.5	80.5	5.3	99.1	0 (0.0)	0.7	2 (3.6)	2.8
34	55 (0.9)	1,242 (19.4)	58.8	81.4	5.4	99.1	1 (1.6)	0.7	2 (3.6)	3.1
35	111 (1.7)	1,187 (18.5)	57	82.2	5.5	99.1	2 (1.6)	0.8	3 (2.7)	3.5
36	179 (2.8)	1,076 (16.8)	54.4	83.9	5.8	99	1 (0.5)	0.9	5 (2.8)	3.9
37	337 (5.3)	897 (14.0)	50	86.7	6.4	99	1 (0.3)	0.9	8 (2.4)	4.3
38	168 (2.6)	560 (8.7)	43	91.9	8.8	99	2 (1.1)	1.0	5 (3.0)	4.8
39	78 (1.2)	392 (6.1)	38.6	94.5	11.2	98.8	0 (0.0)	1.1	7 (9.0)	5.4
40	57 (0.9)	314 (4.9)	32.5	95.6	11.8	98.7	1 (1.7)	1.2	5 (8.8)	6
41	50 (0.8)	257 (4.0)	28.1	96.4	12.5	98.7	0 (0.0)	1.4	3 (6.0)	6.7
42	29 (0.5)	207 (3.2)	25.4	97.2	14	98.6	1 (2.9)	1.5	3 (10.3)	7.5
43	25 (0.4)	178 (2.8)	22.8	97.6	14.6	98.6	1 (3.7)	1.6	6 (24.0)	8.3
44	30 (0.5)	153 (2.4)	17.5	97.9	13.1	98.5	0 (0.0)	1.8	2 (6.7)	9.2
45	33 (0.5)	123 (1.9)	15.8	98.3	14.6	98.5	1 (2.9)	2.0	7 (21.2)	10.3
46	20 (0.3)	90 (1.4)	9.7	98.8	12.2	98.4	0 (0.0)	2.1	1 (5.0)	11.4
47	11 (0.2)	70 (1.1)	8.8	99.1	14.3	98.4	0 (0.0)	2.3	1 (9.1)	12.6
48	14 (0.2)	59 (0.9)	7.9	99.2	15.3	98.4	1 (7.1)	2.6	1 (7.1)	13.9
49	9 (0.1)	45 (0.7)	7	99.4	17.8	98.3	0 (0.0)	2.8	1 (11.1)	15.4
50	8 (0.1)	36 (0.6)	6.1	99.5	19.4	98.3	1 (12.5)	3.1	1 (12.5)	16.9
51	4 (0.1)	28 (0.4)	5.3	99.7	21.4	98.3	0 (0.0)	3.3	0 (0.0)	18.6
52	7 (0.1)	24 (0.4)	5.3	99.7	25	98.3	1 (14.3)	3.7	2 (28.6)	20.4
53	8 (0.1)	17 (0.3)	3.5	99.8	23.5	98.3	1 (12.5)	4.0	3 (37.5)	22.4
54	1 (0.0)	9 (0.1)	0.9	99.8	11.1	98.2	0 (0.0)	4.4	0 (0.0)	24.5
55	1 (0.0)	8 (0.1)	0.9	99.9	12.5	98.2	0 (0.0)	4.8	0 (0.0)	26.7
56	1 (0.0)	7 (0.1)	0.9	99.9	14.3	98.2	0 (0.0)	5.2	0 (0.0)	29
57	1 (0.0)	6 (0.1)	0.9	99.9	16.7	98.2	0 (0.0)	5.7	0 (0.0)	31.4
58	2 (0.0)	5 (0.1)	0.9	99.4	20	98.2	0 (0.0)	6.2	0 (0.0)	34
59	2 (0.0)	3 (0.1)	0.9	99.9	33.3	98.2	0 (0.0)	6.7	0 (0.0)	36.6
64	1 (0.0)	1 (0.0)	0.9	100	100	98.2	0 (0.0)	10.3	1(0.9)	50.7