Annals of Surgery Publish Ahead of Print
DOI:10.1097/SLA.0000000000006544

Retrospective Cohort Study

Development of a Surgery-Specific Comorbidity Score for Use in Administrative Data

Nikhil L Chervu MD MS,[1] Jeff Balian,[1] Arjun Verma BS,[1] Sara Sakowitz MS MPH,[1] Nam Yong Cho BS,[1] Saad Mallick MD,[1] Tara A Russell MD MPH PhD,[2] Peyman Benharash MD[1]

[1]Cardiovascular Outcomes Research Laboratories (CORELAB), David Geffen School of Medicine, University of California, Los Angeles, CA, USA

[2]Division of Colorectal Surgery, Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

Funding/Financial Support: No direct or indirect financial support by extramural sources was received for this research.

Word Count: 3,322/3,500 words

Correspondence:
Peyman Benharash, MD
UCLA Division of Cardiac Surgery
62-249 Center for Health Sciences
Los Angeles, CA, 90095
Tel: (310) 206-6717
Fax: (310) 206-5901
Twitter: @CoreLabUCLA @UCLASurgery
pbenharash@mednet.ucla.edu

Short Running Head - Surgery-Specific Comorbidity Score

Nikhil L Chervu MD MS, Jeff Balian, Arjun Verma BS, Sara Sakowitz MS MPH, Nam Yong Cho BS, Saad Mallick MD, Tara A Russell MD MPH PhD, Peyman Benharash MD

Structured Abstract

Objective

To create a novel comorbidity score tailored for surgical database research.

Summary Background Data

Despite their use in surgical research, the Elixhauser (ECI) and Charlson Comorbidity Indices (CCI) were developed nearly four decades ago utilizing primarily non-surgical cohorts.

Methods

Adults undergoing 62 operations across 14 specialties were queried from the 2019 National Inpatient Sample (NIS) using International Classification of Diseases, 10th Revision (ICD-10) codes. ICD-10 codes for chronic diseases were sorted into Clinical Classifications Software Refined (CCSR) groups. CCSR with non-zero feature importance across four machine learning algorithms predicting in-hospital mortality were used for logistic regression; resultant coefficients were used to calculate the Comorbid Operative Risk Evaluation (CORE) score based on previously validated methodology. Areas under the receiver operating characteristic (AUROC) with 95% Confidence Intervals (CI) were used to compare model performance in predicting in-hospital mortality for the CORE score, ECI, and CCI. Validation was performed using the 2016-2018 NIS, combined 2018-2019 Florida and New York State Inpatient Databases (SID), and 2016-2022 institutional data.

Results

699,155 records from the 2019 NIS were used for model development. The CORE score better predicted in-hospital mortality compared to the ECI within the NIS (0.90, 95%CI:0.90-

0.90 vs 0.84, 95%CI:0.84-0.84), SID (0.91, 95%CI:0.90-0.91 vs 0.86, 95%CI:0.86-0.87), and institutional (0.88, 95%CI:0.87-0.89 vs 0.84, 95%CI:0.83-0.85) databases (all $p<0.001$). Likewise, it outperformed the CCI for the NIS (0.76, 95%CI:0.76-0.76), SID (0.78, 95%CI:0.77-0.78), and institutional (0.62, 95%CI:0.60-0.64) cohorts (all $p<0.001$).

Conclusions

The CORE score may better predict in-hospital mortality after surgery due to comorbid diseases in outcome-based research.

Abstract Word Count: 250/250 words

Introduction

Since their introduction nearly four decades ago, the Charlson Comorbidity Index (CCI) and Elixhauser Comorbidity Index (ECI) have been widely used in healthcare research to quantify the burden of pre-existing conditions.[1–3] Although the superiority of CCI and ECI in specific populations continues to be debated, neither was developed in the context of surgical hospitalizations.[4–6]

In recent years, increasing computational capabilities have enabled healthcare researchers to exploit large databases for highly powered retrospective studies. This has resulted in a significant body of literature focusing on epidemiology, interhospital variation, and disparities in medical care.[7] Unlike their prospective counterparts, retrospective administrative repositories often capture data from medical billing records and lack clinical granularity.[7,8] Thus, it is necessary to develop meaningful ways to incorporate patient comorbidity into risk models has been needed. The first of these tools was the CCI, which was developed using less than 700 breast cancer patients to predict 1-year mortality.[3] Elixhauser et al. subsequently modeled in-hospital mortality using the 1992 California Statewide Inpatient Database. Although they included patients admitted for appendicitis, hernia, and diverticulitis, most were admitted for medical diagnoses such as myocardial infarction, asthma, lower back pain, pneumonia, and diabetes.[1] Application of these comorbidities to surgical cohorts have thus resulted in paradoxical outcomes, with findings suggesting no discernable association with increased mortality risk.[9] Evolving trends in lifestyles, related disease, and the unique nature of the surgical patient necessitate contemporary modalities to appropriately evaluate patients in large healthcare databases.[10,11]

In the present work, we aimed to create a heuristic tool to assess the association between pre-existing conditions and the risk of in-hospital mortality after major operation. We will validate this metric, the Comorbid Operative Risk Evaluation (CORE) score, using nationwide, state-level, and institutional data to improve and validate its performance. This scoring system may represent an improved discriminatory instrument for future risk models and benchmarking across surgical specialties.

Methods

Data Source and Study Population

The CORE score was developed using the 2019 National Inpatient Sample (NIS). Maintained as part of the Healthcare Costs and Utilization Project (HCUP), NIS is the largest, all-payer inpatient database entailing weighted subsets of individual State Inpatient Databases (SID).[12] Data collected by SID contain approximately 97% of all inpatient discharges within a given state.[13] Each record in the NIS and SID can be associated with 40 diagnoses, which are recorded using *International Classification of Disease, 10th Revision* (ICD-10) codes. Relevant codes are captured by medical billing and coding specialists following each hospitalization from physician notes, operative reports, and radiologic or other diagnostic studies. These codes are further grouped by HCUP into over 530 clinical categories from 22 body systems named "Clinical Classifications Software Refined" (CCSR).[14] CCSR have been used in both clinical research and healthcare utilization analyses to objectively define and classify both acute and chronic conditions in administrative data.[15–17] The *elixhauser* and *charlson* Stata commands were used to calculate the ECI and CCI, respectively.

All hospitalization records for adults (≥18 years) undergoing major neurosurgical, otolaryngologic (ENT), endocrine, cardiac, thoracic, acute care surgery (ACS), foregut/bariatric,

hepatopancreatobiliary (HPB), colorectal, urologic, gynecologic, plastic, orthopedic, and vascular operations were identified using relevant ICD-10 codes (Supplemental Table 1, Supplemental Digital Content 1, http://links.lww.com/SLA/F313). The list of major operations was compiled via a thorough review of the available literature for each specialty.[18–32] ACS was not inclusive of trauma due to the difficulties of assessing extent of injury consistently across multiple databases. Pediatric and pregnant patients or those undergoing solid organ transplant were not included. Records missing data for in-hospital mortality, age, sex, or elective case status as coded by each dataset were excluded, as these factors were deemed critical to defining the primary outcome (0.2%; Figure 1). The "SimpleImputer" command from the *sklearn* Python library was used to impute missing values for race, income, primary payer, hospital size, and hospital location/teaching status for machine learning (ML) algorithms.

Model validation was performed using the 2016-2018 NIS, 2018-2019 Florida and New York State Inpatient Databases as well as 2016-2022 University of California, Los Angeles (UCLA) institutional data. By incorporating data from two separate SID captured within the NIS, we could evaluate admissions that were not included in the sampling algorithm used by the NIS. UCLA data included anonymized institutional data from a large quaternary care academic system comprising four hospitals and were obtained from electronic health records. Similar to the NIS and SID, patients were selected based on the presence of the aforementioned ICD-10 procedure codes. Incorporation of these data allowed for further assessment of the validity of the score for both inpatient and outpatient operations, as well as a range of retrospective databases.

Comorbid Conditions and Primary Endpoint

To identify comorbid conditions, all diagnosis codes from admissions for the selected operations were tabulated within the 2019 NIS. Diagnosis codes were then grouped by CCSR group and

further evaluated, whereas those without an associated CCSR were excluded. CCSR groups that were most likely due to non-chronic conditions were then excluded (i.e., CCSR groups with the words “symptom of,” “postoperative,” “postprocedural,” “acute,” or “complication of” in their title). Finally, ICD-10 codes with “acute” in their descriptions were not considered chronic conditions (Supplemental Figure 1, Supplemental Digital Content 2, http://links.lww.com/SLA/F314). In total, 9,811 codes across 325 CCSR were subject to analysis (Supplemental Table 2, Supplemental Digital Content 1, http://links.lww.com/SLA/F313). The primary outcome of the study was in-hospital mortality during the same admission as the operative intervention. This was selected because of its ubiquity as an adverse event across all surgical specialties. Other complications, such as atrial fibrillation, prolonged ventilation, and postoperative transfusion, may be considered less severe or necessary for routine postoperative management after certain cardiac, transplant, and trauma operations. Metrics such as unplanned reoperation and prolonged length of stay are also variable depending on the specialty and operation of interest, thereby limiting the broad applicability of our score.

Model Development and Training

ML has previously been used in healthcare research owing to its superior discrimination and predictive power.[33] Therefore, we used Python to evaluate Random Forest (RF), Gradient Boosting (GBM), eXreme Gradient Boosting (XGBoost), and Multinomial Naïve Bayes (NB) models with 5-fold cross-validation to assist in feature down-selection for logistic regression. Specifically, features with non-zero importance in each of the RF, GBM, XGBoost, and NB models were kept as covariates in the final logistic regression. The 2019 NIS cohort was split into training (80%) and testing (20%) datasets to train and test these models.

Covariates included age, sex, elective case status, race, income quartile, primary payer status, bed size, and hospital location/teaching status, in addition to CCSR codes. Model discrimination was assessed using the area under the receiver operating characteristics (AUROC) and precision recall curves (AUPRC) with 95% confidence intervals (CI) generated by 5-fold cross-validation. Probabilistic estimation accuracy was assessed using Brier scores with 95% CI (Supplemental Figure 2, Supplemental Digital Content 2, http://links.lww.com/SLA/F314).[34] True (TPR) and false positive rates (FPR) as well as sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), balanced accuracy, and reliability scores were also obtained for each model (Supplemental Table 3, Supplemental Digital Content 1, http://links.lww.com/SLA/F313).

Score Development

After ML-assisted feature down-selection, final score development was conducted using parameter estimates derived from logistic regression. This methodology was initially developed by Sullivan et al., who established a mathematical approach to risk score development for multivariable clinical data as part of the Framingham Heart Study.[35] Notably, this algorithm was used by van Walraven to establish a numerical Elixhauser Comorbidity Index.[2] In summary, points are assigned to each CCSR by obtaining parameter estimates from logistic regression. The final point values are calculated by dividing each logistic coefficient by the lowest CCSR coefficient corresponding to the "weakest" (i.e., lowest absolute value) association with in-hospital mortality (Table 1). A final CORE score is then calculated for each patient using the following equation (where b is the absolute value of the weakest estimate):

$$CORE\ Score = 100 * \frac{1}{1 + e^{(intercept + b * point\ total)}}$$

The final score ranges from zero to 100, with zero representing the lowest risk of in-hospital mortality, and 100 being the highest.

We assessed model fit using AUROC with 95% CI using the *roctab* function in Stata. This was compared to the AUROC generated by using the van Walraven modification of the ECI to model mortality.[1,2] Similar methods were used to compare the CORE score to the CCI. In addition to analyzing the NIS, SID, and institutional cohorts overall, additional subgroup analyses by year and specialty were conducted.

Statistical Analyses

Continuous variables are reported as medians with interquartile ranges (IQR). Categorical variables are summarized as frequencies (%). Frequencies and medians were compared using the Pearson's χ^2 and Mann-Whitney U test, respectively. The *roccomp* function in Stata was employed to determine significant differences between AUROC generated by the CORE score and either the ECI or CCI. Due to the large sample size, Cohen's d was used to determine effect size. Cohen's d effect sizes with absolute values ranging from 0.00-0.19 were considered *very small*, 0.20-0.49 *small*, 0.50-0.79 *medium*, and ≥ 0.8 *large* effect size differences.[36] A $p<0.05$ was considered significant. Bonferonni corrections were used for both subgroup analyses due to the large number of comparisons. This led to adjusted *p*-value thresholds of 0.0013 and 0.0004 for by year and by specialty subgroups analyses, respectively. All analyses were conducted using Stata 16.1 and the following Python libraries: *matplotlib* (version 3.7.2), *numpy* (version 1.24.3), *pandas* (version 2.0.3), *sklearn* (version 1.3.0), and *xgboost* (version 2.0.3). Due to the deidentified nature of our data, this study was deemed exempt from full review by the Institutional Review Board at the University of California, Los Angeles (IRB #24-000355).

Results

Model Calibration Data

A total of 699,155 patients were used to develop the model, of which 139,831 (20%) comprised the testing cohort. Baseline characteristics are shown in Table 2. Compared to the training set, the testing group was similar in age (64 [54 - 72] vs 64 [54 - 72], p<0.85, Cohen's d < 0.01), equally female (53.6 vs 53.6%, p=0.88, Cohen's d = <-0.01), and had a comparable proportion of non-elective cases (72.7 vs 72.5%, p=0.14, Cohen's d = <-0.01). There was no difference in median CORE score (23.5 [11.8 - 47.8] vs 23.5 [11.8 - 47.8], p=0.86, Cohen's d < 0.01), ECI (2 [1 - 4] vs 2 [1 - 4], p=0.58, Cohen's d < 0.01), or CCI (1 [0 - 2] vs 1 [0 - 2], p=0.64, Cohen's d < -0.01).

8,272 patients (1.2%) died during their initial hospitalization. Deceased patients had a higher median ECI (6 [4 - 7] vs 2 [1 - 4], p<0.001, Cohen's d = -1.46), CCI (3 [2 - 5] vs 1 [0 - 2], p<0.001, Cohen's d = -1.01), and CORE score (88.4 [70.0 - 96.6] vs 23.4 [12.1 - 46.7], p<0.001, Cohen's d = -1.80), compared to those surviving their initial hospitalization. In addition to the increased effect size between deceased and non-deceased patients, the CORE score (0.90, 95%CI: 0.90 - 0.90) outperformed both the ECI (0.84, 95%CI: 0.83 - 0.84) and CCI (0.76, 95%CI: 0.75 - 0.76) in predicting in-hospital mortality based on AUROC (both p<0.001; Figure 2).

Overall Model Performance

Validation of the model was conducted using the 2016-2018 NIS, 2018-2019 combined Florida and New York SID, and 2016-2022 institutional data. For reference, 944,056 patients were evaluated within the combined SID group, and 34,202 met the inclusion criteria for the institutional cohort. Within the 2016-2018 NIS, the CORE score (85.9 [65.8 - 95.8] vs 21.2 [11.8 - 42.9], p<0.001, Cohen's d = -1.85) demonstrated improved discrimination between deceased

and non-deceased patients as compared to the ECI (6 [4 - 7] vs 2 [1 - 4], p<0.001, Cohen's d = -1.52) and CCI (3 [2 - 5] vs 1 [0 - 2], p<0.001, Cohen's d = -1.06). Improved discrimination between deceased and non-deceased patients was also observed when using the CORE score (90.4 [72.7 - 97.3] vs 23.9 [12.3 - 46.7], p<0.001, Cohen's d = -1.86) as compared to the ECI (6 [5 - 8] vs 2 [1 - 4], p<0.001, Cohen's d = -1.63) and CCI (4 [2 - 6] vs 1 [0 - 2], p<0.001, Cohen's d = -1.15) in the SID cohort. This was also the case for the institutional cohort (CORE score: 89.1 [76.0 - 95.4] vs 35.6 [18.4 - 60.1], Cohen's d = -1.57 vs ECI: 6 [4 - 7] vs 2 [1 - 4], Cohen's d = -1.38 vs CCI 2 [1 - 3] vs 1 [0 - 2], Cohen's d = -0.38; all p<0.001).

The CORE score (0.90, 95%CI: 0.90 - 0.90) yielded a higher AUROC compared to the ECI (0.84, 95%CI: 0.84 - 0.84) or CCI (0.76, 95%CI: 0.76 - 0.76, both p<0.001) for the NIS cohort. This was also the case for the SID population (CORE score: 0.91, 95%CI: 0.90 - 0.91 vs ECI: 0.86, 95%CI: 0.86 - 0.87 and CCI: 0.78, 95%CI: 0.77 - 0.78; both p<0.001). Finally, the ECI (0.84, 95%CI: 0.83 - 0.85) and CCI (0.62, 95%CI: 0.60 - 0.64) did not predict in-hospital mortality as well as the CORE score (0.88, 95%CI: 0.87 - 0.89, both p<0.001) for the institutional cohort (Figure 2).

Subgroup Analysis – Model Performance by Year

A subgroup analysis was conducted to analyze model performance by year. Of note, Bonferroni correction yielded an adjusted p-value threshold of 0.0013. Specifically, the CORE score yielded a higher AUROC compared to the ECI or CCI for each year of the 2016-2018 NIS and 2018-2019 SID (Figures 3a and 3b). For the institutional cohort, the CORE score outperformed the CCI for 2016-2022. However, it only outperformed the ECI from 2019 through 2022. There was no significant difference in the AUROC between the CORE Score and the ECI for institutional data from 2016-2018 and 2022 (Figure 3c).

Subgroup Analysis – Model Performance by Specialty

Additional subgroup analysis was conducted to analyze model performance by specialty with a Bonferroni-corrected *p*-value of 0.0004. Within the NIS cohort, the CORE score outperformed the CCI for all specialties except for ACS (p=0.063). The CORE score likewise predicted in-hospital mortality better than the ECI in 10 out of 14 specialties; the exceptions were endocrine (p=0.002), HPB (p=0.24), gynecology (p=0.59), and plastics (p=0.028; Figure 4a). Compared to ECI, the CORE score yielded a significantly improved AUROC for seven out of the fourteen classified specialties in the SID; endocrine (p=0.33), cardiac (p=0.30), HPB (p=0.71), colorectal (p=0.17), gynecology (p=0.037), plastic (p=0.76), and vascular (p=0.54) operations were the exceptions. It outperformed the CCI for all specialties with the exception of gynecology (p=0.013; Figure 4b). Finally, the CORE score had a significantly higher AUROC compared to the ECI within the institutional cohort for ENT and ACS operations; it performed equivalently for all other specialties. The CORE score outperformed the CCI in six specialties – ENT, cardiac, thoracic, ACS, plastics, and vascular (Figure 4c).

Discussion

Administrative and other retrospective databases allow researchers to increase population size and minimize the selection biases inherent to classical prospective randomized clinical trials.[37] While prospectively collected data can account for disease and project-specific comorbidities, collecting such large-scale data can be resource intensive and subject to biased risk identification.[38] The ECI and CCI were developed in order to address this pitfall of retrospective outcomes research.[1,3] They have likewise been used in benchmarking at the institutional and national levels due to their widespread use.[39] However, these scores were not developed in the

context of operative admissions; studies examining their comparative effectiveness in surgical populations are thus mixed.[4–6]

In the present study, we developed the CORE score to incorporate pre-existing conditions more accurately in surgical database research. It represents the first contemporary comorbidity score specifically designed for multispecialty surgical research using retrospective databases. Prior modifications to the CCI and ECI have yet to account for the baseline differences between surgical and non-surgical patients.[40–42] Patients requiring surgical admission often present electively, with preoperative risk stratification and medical optimization prior to surgery.[11,43] Compared to their non-surgical counterparts, surgical patients are younger, less often frail, and have reduced lengths of stay, thereby minimizing in-hospital complications.[43] These factors, paired with the complex interaction between the stresses of surgery and pre-existing conditions, necessitate surgery-specific methodologies.[10,11,44] In response, individual surgical societies such as the American College of Surgeons and the Society of Thoracic Surgeons have developed preoperative risk stratification tools to aid in patient selection and stratification.[45,46] However, these tools are limited by specialty and the need for granular clinical data not otherwise available in most retrospective databases. Across multiple years, databases, and surgical specialties, the CORE score significantly outperformed both the ECI and CCI in predicting in-hospital mortality after major operation. This is likely due to the use of ML algorithms to assist with feature selection. ML-assisted feature selection has yielded improved predictive models in a wide range of medical research, from chronic obstructive pulmonary disease and breast cancer diagnosis to healthcare costs.[47–49] The POTTER score, developed by Bertsimas et al, is one such tool that has used ML methodology to improve prediction of complications after ACS and trauma surgery.[50] This score is similarly limited by its use of clinical granular data such as labs and recent

diagnosis that restrict its use to purpose-built or institutional databases. However, by incorporating similar techniques with administrative data, we believe that this instrument holds merit as a tool for retrospective health data.

Furthermore, the CORE score provides greater discrimination between deceased and non-deceased patients. In our experience, ECI and CCI medians and confidence intervals between groups can overlap despite the observed statistical significance when dealing with large sample sizes. This is likely due to the inclusion of comorbidities not frequently encountered in surgical populations – congestive heart failure, paralysis, chronic pulmonary disease, renal failure, and liver disease.[2,3] These patients are often deemed too high of a surgical risk to undergo operation. Therefore, the range of possible ECI and CCI values are reduced for surgical patients. When increasing the sample size to tens and hundreds of thousands, however, clinically irrelevant differences can be deemed statistically different. The CORE score increases discrimination by only including comorbidities present in surgical populations and by providing a larger 100-point scale.

Limitations

The present study has several important limitations. As an administrative database, the NIS relies on accurate coding by billing specialists, and may be subject to some error. Furthermore, ICD-10 codes are recorded primarily for financial, and not clinical, purposes. The score is built using the average risk over many people for each condition. We therefore cannot reliably calculate the actual risk of in-hospital mortality for a specific condition for each patient. Finally, rare diagnosis codes that are associated with extremes of risk in mortality may skew the overall score. However, by grouping diagnoses by CCSR and using a large dataset, we attempted to mitigate these risks.

Conclusions

In this large contemporary study, we have established the first comorbidity score for use in administrative database research specifically designed using a surgical cohort. We hope that incorporation of this score in future analyses will allow for more robust adjustment of pre-existing conditions to enhance statistical discrimination. In addition, the increased discriminatory power afforded by a 100-point scale may make subjective analysis of mortality risk easier to determine. Future work applying this score to other studies will allow for continued validation.

Conflict of Interest/Disclosure: PB received fees from AtriCure as a surgical proctor. This manuscript does not discuss any related products or services. Other authors report no conflicts.

Funding/Financial Support: No direct or indirect financial support by extramural sources was received for this research.

Data Sharing Statement: The NIS and SID are publicly available for purchase via the HCUP website. Institutional data is available upon reasonable request with appropriate completion of cross-institutional data sharing agreements. The authors have made a Stata package publicly available within the SSC database. This can be installed using the command *ssc install core*.

Acknowledgements: The authors would like to thank the UCLA Department of Surgery and the Division of Cardiac Surgery, specifically Drs. O. Joe Hines, Timothy Donahue, and Richard Shemin, for their continued support of our research efforts. They would also like to thank Amanda Do, Javier Sanz, and Cenan Pirani for their assistance with obtaining and analysis of institutional data.

References

1. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity Measures for Use with Administrative Data. *Med Care*. 1998;36(1):8-27.

2. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A Modification of the Elixhauser Comorbidity Measures into a Point System for Hospital Death Using Administrative Data. *Med Care*. 2009;47(6):626-633.

3. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8

4. Maron SZ, Neifert SN, Ranson WA, et al. Elixhauser Comorbidity Measure is Superior to Charlson Comorbidity Index In-Predicting Hospital Complications Following Elective Posterior Cervical Decompression and Fusion. *World Neurosurg*. 2020;138:e26-e34. doi:10.1016/j.wneu.2020.01.141

5. Menendez ME, Neuhaus V, van Dijk CN, Ring D. The Elixhauser Comorbidity Method Outperforms the Charlson Index in Predicting Inpatient Death After Orthopaedic Surgery. *Clin Orthop Relat Res*. 2014;472(9):2878-2886. doi:10.1007/s11999-014-3686-7

6. Yang CC, Fong Y, Lin LC, et al. The age-adjusted Charlson comorbidity index is a better predictor of survival in operated lung cancer patients than the Charlson and Elixhauser comorbidity indices. *Eur J Cardiothorac Surg*. 2018;53(1):235-240. doi:10.1093/ejcts/ezx215

7. Adibuzzaman M, DeLaurentis P, Hill J, Benneyworth BD. Big data in healthcare – the promises, challenges and opportunities from a research perspective: A case study with a model database. *AMIA Annu Symp Proc*. 2018;2017:384-392.

8. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: A review of the research literature. *Int J Med Inf*. 2008;77(5):291-304. doi:10.1016/j.ijmedinf.2007.09.001

9. Hewitt J, McCormack C, Tay HS, et al. Prevalence of multimorbidity and its association with outcomes in older emergency general surgical patients: an observational study. *BMJ Open*. 2016;6(3):e010126. doi:10.1136/bmjopen-2015-010126

10. Neumayer L, Ghalyaie N. Chapter 10 - Principles of Preoperative and Operative Surgery. In: Townsend CM, Beauchamp RD, Evers BM, Mattox KL, eds. *Sabiston Textbook of Surgery - The Biological Basis of Modern Surgical Practice*. 20th ed. Elsevier; 2017:2176.

11. Fowler AJ, Wahedally MAH, Abbott TEF, Prowle JR, Cromwell DA, Pearse RM. Long-term disease interactions amongst surgical patients: a population cohort study. *BJA Br J Anaesth*. 2023;131(2):407-417. doi:10.1016/j.bja.2023.04.041

12. HCUP-US NIS Overview. Accessed April 11, 2023. https://hcup-us.ahrq.gov/nisoverview.jsp

13. HCUP-US SID Overview. Accessed May 17, 2024. https://hcup-us.ahrq.gov/sidoverview.jsp

14. Clinical Classifications Software Refined (CCSR). Accessed May 17, 2024. https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp

15. Johnston S, Jha A, Roy S, Pollack E. Surgical Complication Risk Factor Identification Using High-Dimensional Hospital Data: An Illustrative Example in Hemostasis-Related Complications. *Clin Outcomes Res*. 2022;14:683-689. doi:10.2147/CEOR.S380004

16. Weiss AJ, Jiang HJ. Overview of Clinical Conditions With Frequent and Costly Hospital Readmissions by Payer, 2018. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Agency for Healthcare Research and Quality (US); 2006. Accessed April 30, 2024. http://www.ncbi.nlm.nih.gov/books/NBK573265/

17. McDermott KW, Roemer M. Most Frequent Principal Diagnoses for Inpatient Stays in U.S. Hospitals, 2018. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Agency for Healthcare Research and Quality (US); 2006. Accessed April 30, 2024. http://www.ncbi.nlm.nih.gov/books/NBK573113/

18. McCutcheon BA, Hirshman BR, Gabel BC, et al. Impact of neurosurgeon specialization on patient outcomes for intracranial and spinal surgery: a retrospective analysis of the Nationwide Inpatient Sample 1998–2009. *J Neurosurg*. 2017;128(5):1578-1588. doi:10.3171/2016.4.JNS152332

19. Crippen MM, Barinsky GL, Reddy RK, et al. The Impact of Duty-Hour Restrictions on Complication Rates Following Major Head and Neck Procedures. *The Laryngoscope*. 2018;128(12):2804-2810. doi:10.1002/lary.27338

20. Ebrahimian S, Chervu N, Hadaya J, et al. National outcomes of expedited discharge following esophagectomy for malignancy. *PLOS ONE*. 2024;19(2):e0297470. doi:10.1371/journal.pone.0297470

21. Tran Z, Chervu N, Williamson C, et al. The Impact of Expedited Discharge on 30-Day Readmission After Lung Resection: A National Study. *Ann Thorac Surg*. 2022;113(4):1274-1281. doi:10.1016/j.athoracsur.2021.04.009

22. Seo YJ, Christian-Miller N, Aguayo E, Sanaiha Y, Benharash P, Yanagawa J. National Use and Short-term Outcomes of Video and Robot-Assisted Thoracoscopic Thymectomies. *Ann Thorac Surg*. 2022;113(1):230-236. doi:10.1016/j.athoracsur.2021.02.003

23. Sakowitz S, Bakhtiyar SS, Gao Z, et al. Interhospital Transfer for Emergency General Surgery: A Contemporary National Analysis. *Am Surg*. Published online April 3, 2024:00031348241244642. doi:10.1177/00031348241244642

24. Till BM, Mack SJ, Whitehorn G, et al. The Epidemiology of Surgically Managed Hiatal Hernia: A Nine Year Review of National Trends. *Foregut*. 2023;3(1):20-28. doi:10.1177/26345161221137089

25. Ng AP, Sanaiha Y, Bakhtiyar SS, Ebrahimian S, Branche C, Benharash P. National analysis of cost disparities in robotic-assisted versus laparoscopic abdominal operations. *Surgery*. 2023;173(6):1340-1345. doi:10.1016/j.surg.2023.02.016

26. Mabeza RM, Mao Y, Maynard K, Lee C, Benharash P, Yetasook A. Bariatric surgery outcomes in geriatric patients: a contemporary, nationwide analysis. *Surg Obes Relat Dis*. 2022;18(8):1005-1011. doi:10.1016/j.soard.2022.04.014

27. Williamson CG, Ebrahimian S, Sakowitz S, et al. Race, Insurance, and Sex-Based Disparities in Access to High-Volume Centers for Pancreatectomy. *Ann Surg Oncol*. 2023;30(5):3002-3010. doi:10.1245/s10434-022-13032-8

28. Madrigal J, Hadaya J, Lee C, Tran Z, Benharash P. Association of Frailty with Perioperative Outcomes Following Hepatic Resection: A National Study. *J Am Med Dir Assoc*. 2022;23(4):684-689.e1. doi:10.1016/j.jamda.2022.02.004

29. Mazzone E, Mistretta FA, Knipper S, et al. Contemporary National Assessment of Robot-Assisted Surgery Rates and Total Hospital Charges for Major Surgical Uro-Oncological Procedures in the United States. *J Endourol*. 2019;33(6):438-447. doi:10.1089/end.2018.0840

30. Chan JK, Gardner AB, Mann AK, Kapp DS. Hospital-acquired conditions after surgery for gynecologic cancer — An analysis of 82,304 patients. *Gynecol Oncol*. 2018;150(3):515-520. doi:10.1016/j.ygyno.2018.07.009

31. Sayegh F, Warburton A, Sanati-Mehrizy P, Burish NM, Taub PJ. Determining the Factors Associated With 30- and 90-day Readmissions Following Plastic Surgery Within the National Readmissions Database. *Plast Reconstr Surg – Glob Open*. 2020;8(9S):132. doi:10.1097/01.GOX.0000721012.03765.82

32. Choi YS, Kim TW, Chang MJ, Kang SB, Chang CB. Enhanced recovery after surgery for major orthopedic surgery: a narrative review. *Knee Surg Relat Res*. 2022;34(1):8. doi:10.1186/s43019-022-00137-3

33. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. *JAMA*. 2018;319(13):1317-1318. doi:10.1001/jama.2017.18391

34. Verma A, Balian J, Hadaya J, et al. Machine Learning Based Prediction of Postoperative Pancreatic Fistula Following Pancreaticoduodenectomy. *Ann Surg*.:10.1097/SLA.0000000000006123. doi:10.1097/SLA.0000000000006123

35. Sullivan LM, Massaro JM, D'Agostino Sr. RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*. 2004;23(10):1631-1660. doi:10.1002/sim.1742

36. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Routledge; 2013.

37. Guller U. Surgical Outcomes Research Based on Administrative Data: Inferior or Complementary to Prospective Randomized Clinical Trials? *World J Surg*. 2006;30(3):255-266. doi:10.1007/s00268-005-0156-0

38. Talari K, Goyal M. Retrospective Studies – Utility and Caveats. *J R Coll Physicians Edinb*. 2020;50(4):398-402. doi:10.4997/jrcpe.2020.409

39. Xiang J, Durance PW, Griffes LC, Chen Y, Bakshi RR. Measuring case severity: a novel tool for benchmarking and clinical documentation improvement. *BMC Health Serv Res*. 2022;22(1):513. doi:10.1186/s12913-022-07935-1

40. Holman CDJ, Preen DB, Baynham NJ, Finn JC, Semmens JB. A multipurpose comorbidity scoring system performed better than the Charlson index. *J Clin Epidemiol*. 2005;58(10):1006-1014. doi:10.1016/j.jclinepi.2005.01.020

41. Quan H, Li B, Couris CM, et al. Updating and Validating the Charlson Comorbidity Index and Score for Risk Adjustment in Hospital Discharge Abstracts Using Data From 6 Countries. *Am J Epidemiol*. 2011;173(6):676-682. doi:10.1093/aje/kwq433

42. Klabunde CN, Potosky AL, Legler JM, Warren JL. Development of a comorbidity index using physician claims data. *J Clin Epidemiol*. 2000;53(12):1258-1267. doi:10.1016/S0895-4356(00)00256-0

43. Boeker EB, Ram K, Klopotowska JE, et al. An individual patient data meta-analysis on factors associated with adverse drug events in surgical and non-surgical inpatients. *Br J Clin Pharmacol*. 2015;79(4):548-557. doi:10.1111/bcp.12504

44. Wang J, Chen K, Li X, et al. Postoperative adverse events in patients with diabetes undergoing orthopedic and general surgery. *Medicine (Baltimore)*. 2019;98(14):e15089. doi:10.1097/MD.0000000000015089

45. Bilimoria KY, Liu Y, Paruch JL, et al. Development and Evaluation of the Universal ACS NSQIP Surgical Risk Calculator: A Decision Aid and Informed Consent Tool for Patients and Surgeons. *J Am Coll Surg*. 2013;217(5):833-842.e3. doi:10.1016/j.jamcollsurg.2013.07.385

46. Raphael CE, Sabbagh AE, Rihal C. CHAPTER 7 - Transcatheter aortic valve implantation: Work up and indications for intervention. In: Rihal CS, Raphael CE, eds. *Handbook of Structural Heart Interventions*. Elsevier; 2021:81-87. doi:10.1016/B978-0-323-67278-8.00016-8

47. Moll M, Qiao D, Regan EA, et al. Machine Learning and Prediction of All-Cause Mortality in COPD. *Chest*. 2020;158(3):952-964. doi:10.1016/j.chest.2020.02.079

48. Panay B, Baloian N, Pino JA, Peñafiel S, Sanson H, Bersano N. Feature Selection for Health Care Costs Prediction Using Weighted Evidential Regression. *Sensors*. 2020;20(16):4392. doi:10.3390/s20164392

49. Sharma A, Mishra PK. Performance analysis of machine learning based optimized feature selection approaches for breast cancer diagnosis. *Int J Inf Technol*. 2022;14(4):1949-1960. doi:10.1007/s41870-021-00671-5

50. Bertsimas D, Dunn J, Velmahos GC, Kaafarani HMA. Surgical Risk Is Not Linear: Derivation and Validation of a Novel, User-friendly, and Machine-learning-based Predictive OpTimal Trees in Emergency Surgery Risk (POTTER) Calculator. *Ann Surg*. 2018;268(4):574. doi:10.1097/SLA.0000000000002956

Figure 1. CONSORT diagram of patients included from the 2019 National Inpatient Sample to build the Comorbid Operative Risk Evaluation Score

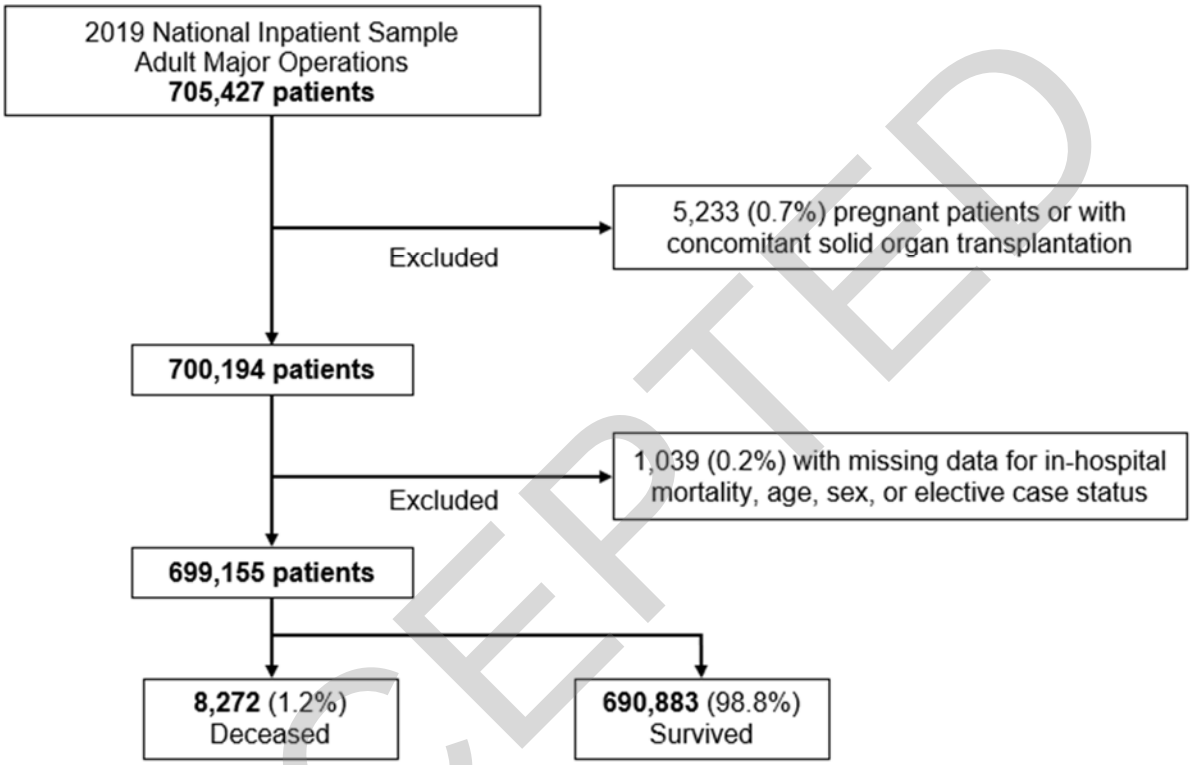

Figure 2. Area Under the Receiver Operator Characteristics (AUROC) with 95% confidence intervals between the Comorbid Operative Risk Evaluation (CORE) Score, Elixhauser Comorbidity Index (ECI), and Charlson Comorbidity Index (CCI) in predicting in-hospital mortality for the National Inpatient Sample (NIS), combined Florida and New York State Inpatient Database (SID), and institutional data

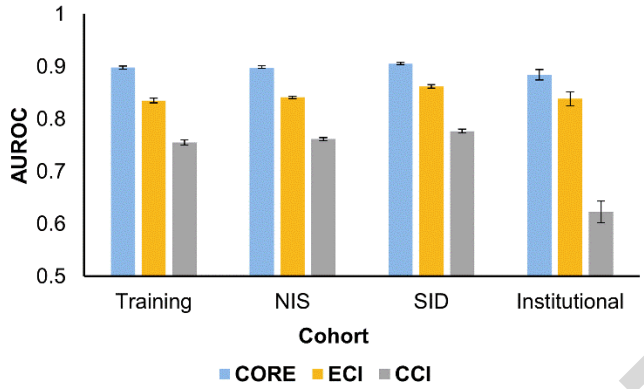

Figure 3. Comparison of Area Under the Receiver Operator Characteristics (AUROC) with 95% confidence intervals between the Comorbid Operative Risk Evaluation (CORE) Score, Elixhauser Comorbidity Index (ECI), and Charlson Comorbidity Index (CCI) by year in predicting in-hospital mortality for the a) 2016-2018 National Inpatient Sample (NIS), b) 2018-2019 combined Florida and New York State Inpatient Database (SID), and c) 2016-2022 institutional data

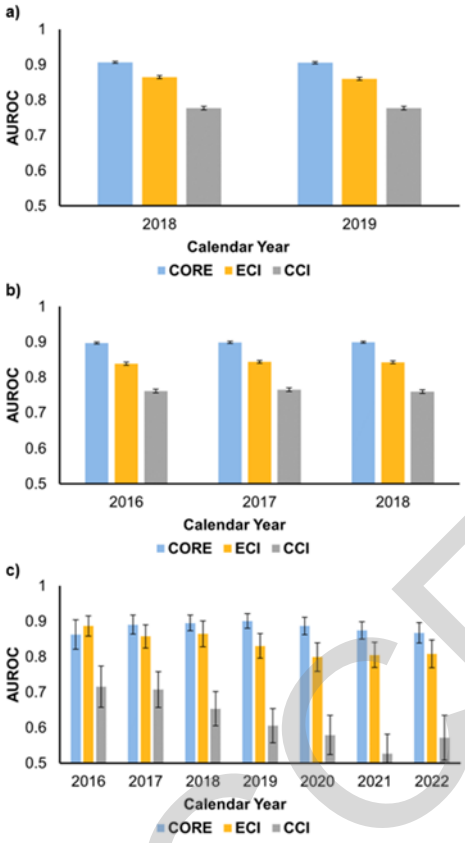

Figure 4. Comparison of Area Under the Receiver Operator Characteristics (AUROC) with 95% confidence intervals between the Comorbid Operative Risk Evaluation (CORE) Score, Elixhauser Comorbidity Index (ECI), and Charlson Comorbidity Index (CCI) by specialty in predicting in-hospital mortality for the a) 2016-2018 National Inpatient Sample (NIS), b) 2018-2019 combined Florida and New York State Inpatient Database (SID), and c) 2016-2022 institutional data; ENT, otolaryngology; ACS, acute care surgery; HPB, hepatopancreatobiliary

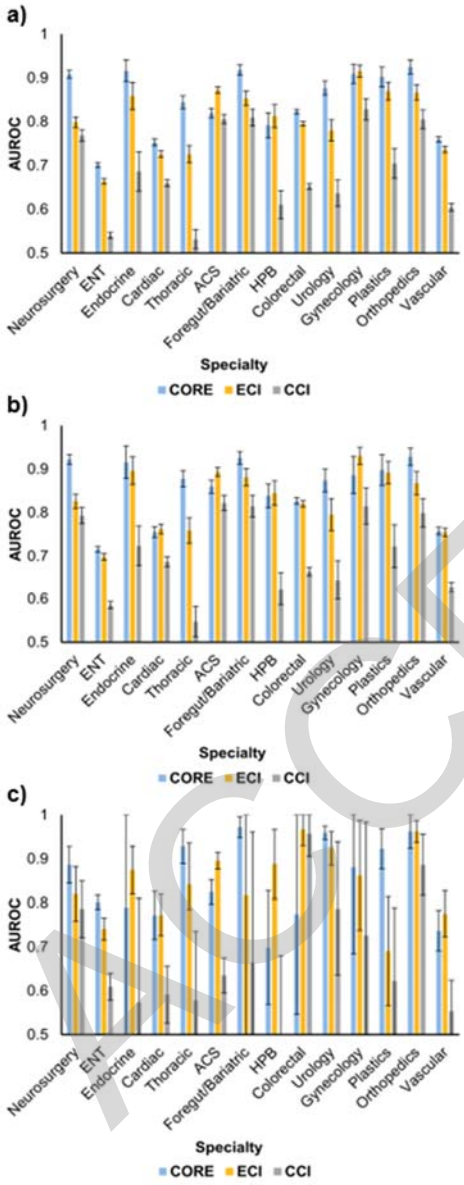

Table 1. Logistic regression coefficients and point totals of Clinical Classifications Software Refined (CCSR) selected by machine learning-assisted feature down-selection

CCSR	Logistic Coefficient	Point Total	Rounded Total	CCSR	Logistic Coefficient	Point Total	Rounded Total
BLD001	-0.4140742	-16.118894	-16	GEN003	0.02616477	1.0185302	1
BLD007	-0.4339951	-16.894366	-17	GEN004	-1.0403581	-40.49859	-40
CIR003	0.41216456	16.044556	16	GEN006	0.05112456	1.9901538	2
CIR007	0.07422739	2.8894903	3	GEN007	0.14358861	5.5895526	6
CIR008	-0.6356478	-24.744213	-25	GEN009	-0.0599012	-2.3318063	-2
CIR011	-0.3359079	-13.076073	-13	GEN021	0.54944763	21.38865	21
CIR019	-0.4282873	-16.672177	-17	INF003	-0.1941078	-7.5561394	-8
CIR020	-0.6059602	-23.588544	-24	MAL001	0.34800999	13.547176	14
CIR026	-0.0422188	-1.6434755	-2	MBD017	-0.6529413	-25.417405	-25
CIR027	-0.3354948	-13.059991	-13	MBD019	-0.3875139	-15.084967	-15
CIR028	-0.4386213	-17.074452	-17	MBD021	-0.4043921	-15.741995	-16
CIR030	-0.5700281	-22.189797	-22	MUS002	-0.7716762	-30.039465	-30
CIR032	-0.4479241	-17.436587	-17	MUS006	1.39251333	54.207132	54
CIR036	-0.1300005	-5.0606017	-5	MUS011	0.78950704	30.733574	31
DIG004	0.34137531	13.288905	13	MUS014	-0.3524071	-13.718347	-14
DIG006	-0.5587775	-21.751838	-22	MUS023	0.29991283	11.674872	12
DIG007	-0.4152166	-16.163367	-16	***MUS029***	***0.02568875***	***1***	***1***
DIG009	-0.3799268	-14.78962	-15	NEO002	0.3197803	12.448264	12
DIG010	-0.2848255	-11.087557	-11	NEO015	0.62129091	24.185333	24
DIG012	-1.1192772	-43.570718	-44	NEO022	0.68063511	26.495457	26
DIG013	-0.4766364	-18.554286	-19	NEO039	0.89349186	34.781449	35
DIG014	-0.369295	-14.375749	-14	NEO043	0.36839223	14.340607	14
DIG016	-0.718447	-27.96738	-28	NEO044	0.16837199	6.554309	7
DIG017	-1.8160148	-70.693005	-71	NEO045	0.81104885	31.572145	32
DIG021	-0.5101593	-19.85925	-20	NEO051	0.37479276	14.589764	15
DIG022	-0.3844221	-14.964612	-15	NEO070	0.46763491	18.203881	18
DIG025	-0.2300124	-8.9538187	-9	NEO073	1.01500978	39.511844	40
END003	-0.3073005	-11.962454	-12	NEO074	-0.1447766	-5.635799	-6
END008	-0.6377707	-24.826851	-25	NVS008	-0.9038077	-35.183019	-35
END009	0.44771479	17.42844	17	NVS011	-0.7202688	-28.0383	-28
END010	0.12113647	4.715546	5	NVS014	-0.2493991	-9.708497	-10
END011	-0.8058898	-31.371317	-31	NVS016	0.40443441	15.743641	16

Table 2. Patient, clinical, hospital characteristics testing and training cohorts derived from the 2019 National Inpatient Sample used to develop the Comorbid Operative Risk Evaluation (CORE) Score; IQR, interquartile range; ECI, Elixhauser Comorbidity Index; CCI, Charlson Comorbidity Index; ENT, ear nose and throat; ACS, acute care surgery; HPB, hepatopancreatobiliary

	Testing	Training	p-value	Cohen's d
Age (years, median [IQR])	64 [54 - 72]	64 [54 - 72]	0.85	< 0.01
Female (%)	75,003 (53.6)	299,887 (53.6)	0.88	< - 0.01
Elective Case Status (%)	101,637 (72.7)	405,431 (72.5)	0.14	< -0.01
In-Hospital Mortality (%)	1,689 (1.2)	6,583 91.2)	0.34	< -0.01
Scores (median [IQR])				
CORE score	23.5 [11.8 - 47.8]	23.5 [11.8 - 47.8]	0.86	< 0.01
ECI	2 [1 - 4]	2 [1 - 4]	0.58	< 0.01
CCI	1 [0 - 2]	1 [0 - 2]	0.64	< - 0.01
Race (%)			0.20	< 0.01
White	102,343 (73.2)	409,269 (73.2)		
Black	15,603 (11.2)	61,687 (11.0)		
Hispanic	11,635 (8.3)	46,542 (8.3)		
Asian/Pacific Islander/Other	6,900 (4.9)	28,265 (5.1)		
Income Quartile (%)			0.76	< - 0.01
76th - 100th	31,324 (22.4)	125,586 (22.5)		
51st - 75th	36,271 (25.9)	144,180 (25.8)		
26th - 50th	35,013 (25.0)	140,106 (25.1)		
0 - 25th	35,205 (25.2)	141,265 (25.3)		
Primary Payer (%)			0.94	< 0.01
Private	49,957 (35.7)	200,016 (35.8)		
Medicare	68,241 (48.8)	272,711 (48.8)		
Medicaid	13,005 (9.3)	51,844 (9.3)		
Uninsured/Other	8,470 (6.1)	34,076 (6.1)		
Specialty (%)				
Neurosurgery	18,970 (13.6)	75,550 (13.5)	0.56	< - 0.01
ENT	4,823 (3.5)	19,674 (3.5)	0.21	< 0.01
Endocrine	1,222 (0.9)	5,120 (0.9)	0.14	< 0.01
Cardiac	10,833 (7.8)	43,646 (7.8)	0.48	< 0.01
Thoracic	3,010 (2.2)	12,207 (2.2)	0.49	< 0.01
ACS	11,449 (8.2)	45,737 (8.2)	0.90	< - 0.01
Foregut/Bariatric	10,826 (7.7)	43,055 (7.7)	0.58	< - 0.01
HPB	1,371 (1.0)	5,373 (1.0)	0.50	< - 0.01
Colorectal	11,222 (8.0)	45,948 (8.2)	0.02	0.01
Urology	5,341 (3.8)	20,977 (3.8)	0.22	< - 0.01
Gynecology	6,582 (4.7)	26,209 (4.7)	0.74	< - 0.01
Plastics	1,653 (1.2)	6,762 (1.2)	0.41	< 0.01
Orthopedics	43,510 (31.1)	173,738 (31.1)	0.70	< - 0.01
Vascular	13,886 (9.9)	55,056 (9.8)	0.33	< - 0.01