Differentiating Urgent from Elective Cases Matters in Minority Populations: Developing an Ordinal "Desirability of Outcome Ranking" to Increase Granularity and Sensitivity of Surgical Outcomes Assessment

Michael A Jacobs, MS, Susanne Schmidt, PhD, Daniel E Hall, MD, MDiv, MHSc, FACS, Karyn B Stitzenberg, MD, MPH, FACS, Lillian S Kao, MD, MS, FACS, Chen-Pin Wang, PhD, Laura S Manuel, BS, Paula K Shireman, MD, MS, MBA, FACS

BACKGROUND: Surgical analyses often focus on single or binary outcomes; we developed an ordinal Desirability of Outcome Ranking (DOOR) for surgery to increase granularity and sensitivity of surgical outcome assessments. Many studies also combine elective and urgent procedures for risk adjustment. We used DOOR to examine complex associations of race/ethnicity and presentation acuity.

STUDY DESIGN: NSQIP (2013 to 2019) cohort study assessing DOOR outcomes across race/ethnicity groups risk-adjusted for frailty, operative stress, preoperative acute serious conditions, and elective, urgent, and emergent cases.

RESULTS: The cohort included 1,597,199 elective, 340,350 urgent, and 185,073 emergent cases with patient mean age of 60.0 ± 15.8, and 56.4% of the surgeries were performed on female patients. Minority race/ethnicity groups had increased odds of presenting with preoperative acute serious conditions (adjusted odds ratio [aORs] range 1.22 to 1.74), urgent (aOR range 1.04 to 2.21), and emergent (aOR range 1.15 to 2.18) surgeries vs the White group. Black (aOR range 1.23 to 1.34) and Native (aOR range 1.07 to 1.17) groups had increased odds of higher/worse DOOR outcomes; however, the Hispanic group had increased odds of higher/worse DOOR (aOR 1.11, CI 1.10 to 1.13), but decreased odds (aORs range 0.94 to 0.96) after adjusting for case status; the Asian group had better outcomes vs the White group. DOOR outcomes improved in minority groups when using elective vs elective/urgent cases as the reference group.

CONCLUSIONS: NSQIP surgical DOOR is a new method to assess outcomes and reveals a complex interplay between race/ethnicity and presentation acuity. Combining elective and urgent cases in risk adjustment may penalize hospitals serving a higher proportion of minority populations. DOOR can be used to improve detection of health disparities and serves as a roadmap for the development of other ordinal surgical outcomes measures. Improving surgical outcomes should focus on decreasing preoperative acute serious conditions and urgent and emergent surgeries, possibly by improving access to care, especially for minority populations. (J Am Coll Surg 2023;237: 545–555. Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government.)

Disclosure Information: Dr Hall receives consulting fees from FutureAssure, LLC. Dr Kao receives royalty payments from Springer, Wolters Kluwer, and McGraw-Hill.

Support: This research was supported by grant U01TR002393 (Jacobs, Hall, Stitzenberg, Kao, Wang, Schmidt, Manuel, and Shireman), from the National Center for Advancing Translational Sciences and the Office of the Director, NIH and Clinical Translational Science Awards UL1TR002489 (NC Translational and Clinical Science [NC TraCS], University of North Carolina), UL1TR001857 (University of Pittsburgh) and UL1TR002645 (University of Texas Health San Antonio) from the National Center for Advancing Translational Sciences, NIH and P30AG044271 from the National Institute for Aging, NIH (Shireman). Dr Hall received grants from the National Institutes of Health and Veterans Health Administration Office of Research and Development outside the scope of this work (I01HX003215, I01HX003095, 1I01HX003322, HX003201-01). Dr Stitzenberg and Ms Manuel received grant funding from the National Institutes of Health. Dr

https://doi.org/10.1097/XCS.0000000000000776
ISSN 1879-1190/22

Abbreviations and Acronyms

ACS	=	American College of Surgeons
aOR	=	adjusted odds ratio
DOOR	=	desirability of outcome ranking
OSS	=	operative stress score
PASC	=	preoperative acute serious conditions
PUF	=	Participant Use File
RAI	=	risk analysis index

Surgical outcomes are strongly associated with the acuity of presentation.[1-5] Presentation acuity has been measured by presenting with preoperative acute serious conditions (PASC)[1-4] and elective, urgent, and emergent case status.[5-7] Emergency surgeries are associated with worse outcomes[8-10] and are independent risk factors for mortality and postoperative complications.[8] Urgent surgeries are more time sensitive than strictly planned/elective cases. Urgent cases usually occur during unplanned hospitalizations after a failed trial of conservative medical management and/or optimization before an unplanned surgery.[1] Similar to emergency surgeries, urgent cases have worse outcomes compared to elective surgeries.[5-7] However, common approaches to risk adjustment merge the urgent and elective categories into a composite reference group compared to emergent cases.[8-10] This approach to risk adjustment may have an even greater impact on safety-net hospitals treating vulnerable populations compared to hospitals with more resources, because minority populations have higher rates of both emergent and urgent surgeries.[3,11-13]

Additionally, most studies focus on a single outcome after surgery, which may not adequately capture health disparities related to race/ethnicity used as a social construct.[14] Composite variables (ie those combining multiple outcomes) such as textbook outcomes[15,16] mitigate some of these weaknesses[17]; however, textbook outcomes suffers from reduced power as a binary outcome and treats all negative outcomes as having similar severity.[18] Desirability of outcome ranking (DOOR) methodology uses a single, ordinal rank to assess treatment or intervention outcomes. Ranking the severity of multiple poor outcomes is a more granular way to assess and compare surgical results compared to single or binary outcomes[16,19] and may improve detection of complex, multivalent risks,[20,21] especially when evaluating the impact of social determinants of health variables on outcomes. DOOR was originally developed to compare antibiotic effectiveness in clinical trials[22-24] but has not been used for surgical outcomes.

Therefore, we developed a surgical DOOR[22-24] with higher scores representing worse/less desirable outcomes to examine the association of elective, urgent, and emergent case status on outcomes using the American College of Surgeons (ACS) NSQIP. We used high-quality, nurse-abstracted NSQIP data[25,26] to mitigate the known limitations of using administrative databases that lack detailed data on patient risk factors and outcomes,[26,27] including variations in assigning ICD9/10 codes across institutions.[25,28,29] We hypothesize that (1) presentation acuity, measured by PASC[2,4,12] and case status, will be higher in racial/ethnic minority populations; (2) increased presentation acuity will be associated with increased/worse DOOR outcomes; (3) worse DOOR outcomes across minority populations will be differentially associated with presentation acuity; and (4) worse DOOR outcomes across minority groups will be altered based on the choice of case status reference group (elective vs combining elective and urgent cases). Our goals were to assess associations using the more granular DOOR and evaluate the effect of combining elective and urgent cases on risk adjustment across race/ethnicity groups.

Shireman reported receiving grants from the National Institutes of Health and Veterans Health Administration and salary support from Texas A&M University, South Texas Veterans Health Care System and the University of Texas Health San Antonio during the conduct of the study. Dr Schmidt reports receiving grants from the National Institutes of Health and Veterans Health Administration.

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. The American College of Surgeons NSQIP and participating hospitals are the source of the data used herein; they have not been verified and are not responsible for the statistical validity of the data analysis or the conclusions derived by the authors.

Received December 1, 2022; Revised March 1, 2023; Accepted March 1, 2023.

From the Department of Surgery, University of Texas Health San Antonio, San Antonio, TX (Jacobs, Shireman); Department of Population Health Sciences, University of Texas Health San Antonio, San Antonio, TX (Schmidt, Wang); Center for Health Equity Research and Promotion, and Geriatric Research Education and Clinical Center, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA (Hall); Department of Surgery, University of Pittsburgh, Pittsburgh, PA (Hall); Wolff Center, UPMC, Pittsburgh, PA (Hall); Department of Surgery, University of North Carolina, Chapel Hill, NC (Stitzenberg); Department of Surgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX (Kao); UT Health Physicians Business Intelligence and Data Analytics, University of Texas Health San Antonio, San Antonio, TX (Manuel); University Health, San Antonio, TX (Shireman); and Departments of Primary Care & Rural Medicine and Medical Physiology, School of Medicine, Texas A&M Health, Bryan, TX (Shireman).

Correspondence address: Paula K Shireman, MD, MS, MBA, Office of the Dean, School of Medicine, Texas A&M Health, 8447 Riverside Pkwy, Bryan, TX 77807. email: Shireman@tamu.edu

Supplemental digital content is available for this article.

METHODS

Study population and data

This retrospective cohort study included inpatient procedures in the 2013 to 2019 ACS NSQIP Participant Use File (PUF), following STROBE[30] reporting guidelines. Patients not admitted from home, transferred to another acute care hospital, or discharged against medical advice were excluded. This study was deemed exempt by the University of Texas Health San Antonio Institutional Review Board.

Estimating patient frailty/premorbid conditions

The Risk Analysis Index (RAI) assesses frailty using NSQIP variables. RAI has been validated in multiple datasets[31-33] and renders a score ranging from 0 to 81 categorized as robust (≤20), normal (21 to 29), frail (30 to 39), and very frail (≥40).[31] RAI is used as a single variable estimate of patient-level variability that overcomes barriers to model fit encountered by less parsimonious models.[4,34-37]

Presentation acuity: preoperative acute serious conditions and case status

Patients presenting with PASC were defined using 6 NSQIP "present at time of surgery" variables and NSQIP variables defining preoperative acute renal failure without or combined with dialysis required within 2 weeks before surgery, as previously described[4] (see **Supplemental Digital Content 1**, http://links.lww.com/JACS/A276 for PASC variables).

Case status was determined from NSQIP variables with urgent cases being defined as neither elective nor emergent, as determined by "no" responses to the ELECTSURG and EMERGNCY variables.[5] NSQIP defines elective and emergent variables to exclude urgent cases.[38]

Expanded operative stress score assignment

The operative stress score (OSS) estimates surgical-induced physiologic stress of procedures across surgical specialties based on CPT codes by assigning a score ranging from 1 to 5, with 1 and 5 representing very low and very high physiologic stress, respectively. Similar to the RAI, the OSS is a single-variable estimate of procedure-level variability, overcoming barriers to model fit encountered with less parsimonious models. We used the expanded OSS[4] with 2,343 CPT codes, which includes gynecologic and other procedures omitted from the original OSS. The expanded OSS includes a substantially greater proportion of the cases included in NSQIP, especially among female populations.[37] After excluding cases without an expanded OSS assigned to the principal CPT code, OSS was assigned using the highest score for all available CPT codes within each case.[4]

Race and ethnicity

Patient race and ethnicity data, used as a social construct,[14] were derived from NSQIP variables of RACE_NEW and ETHNICITY_HISPANIC. We categorized race/ethnicity as (1) Asian, non-Hispanic; (2) Black, non-Hispanic; (3) Hispanic of any race; (4) Native,[39] including American Indian, Alaska Native, Native Hawaiian, or Pacific Islander, all non-Hispanic; and (5) White, non-Hispanic. Although different terminologies exist, we used "Native" as the group term based on recommendations from the Smithsonian National Museum of the American Indian.[40]

Surgical desirability of outcome ranking

The 21 NSQIP variables for 30-day complications, unplanned readmissions, and mortality were ranked by severity (**Supplemental Digital Content 2**, http://links.lww.com/JACS/A276) using a modified Delphi process.[41,42] Mortality 30 days from the date of surgery was defined using the DOPERTOD variable, with values ranging from 0 to 30 counted as "Yes" and -99 counted as "No." Authors from multiple surgical specialties (DEH, KBS, LSK, and PKS) independently rated the severity of each surgical complication. DOOR rankings were assigned using a scale of 1 representing no complications/most desirable outcome and ranks of 2 to 6 were assigned to progressively worse/less desirable outcomes. Rankings were summarized and shared with each subsequent round of ranking. Disagreements between rankings were discussed with each round to build consensus. Rankings occurred during 3 rounds to reach a final, unanimous consensus.

Study outcomes

Our main outcomes were the association of presentation acuity (measured by PASC and elective, urgent, and emergent case status) and race/ethnicity with DOOR outcomes, adjusted for RAI and OSS. Secondary outcomes included association of race/ethnicity with PASC and undergoing urgent and emergent cases and subgroup analyses for each case status (elective, urgent, and emergent) of DOOR outcomes. Sensitivity analyses assessed the differences in odds of undergoing emergent cases and DOOR outcomes between models using elective vs combined elective/urgent cases as the reference group.

Statistical analysis

Categorical data were summarized using counts and percentages and continuous data using mean ± SD. Chi-square tests and F-tests were used to test for differences between groups for categorical and continuous variables.

We assessed the association between race/ethnicity and case status with DOOR by using ordinal logistic (proportional odds) regression models. DOOR analyses were performed on the elective, urgent, and emergent subgroups. Secondary analyses used ordinal logistic regression models to compare the association of race/ethnicity on presenting with PASC and undergoing urgent or emergent surgeries vs elective, with a sensitivity analysis of emergent cases using elective plus urgent cases as the reference group. These analyses used nested models to examine the differing effects of various risk adjustment factors. All p values were assessed using a 2-tailed significance of $p < 0.05$ and confidence intervals are reported at the 95% level. Analyses were performed using R 4.2.0 (R Project for Statistical Computing).

RESULTS

Population characteristics

The cohort included 2,122,622 inpatient cases contained in the 2013 to 2019 NSQIP PUF (**Supplemental Digital Content 3**, http://links.lww.com/JACS/A276). Most patients identified as White, non-Hispanic (74.9%, **Supplemental Digital Content 4**, http://links.lww.com/JACS/A276) followed by Black, non-Hispanic (12.0%); Hispanic, any race (8.7%); Asian, non-Hispanic (3.5%); and Native, non-Hispanic (0.9%). Patients had a mean ± SD age of 60.0 ± 15.8 years and were primarily female (56.4%). Cases were most commonly elective (75.2%), followed by urgent (16.0%) and emergent (8.7%). Higher proportions of robust patients (**Supplemental Digital Content 5**, http://links.lww.com/JACS/A276) were present in the Hispanic, Black, and Native groups (68.7%, 65.4% and 63.4%, respectively) compared to Asian (57.4%) and White (50.2%) groups. Patients identified as White had the highest odds of undergoing higher OSS procedures followed by Black, then Asian, with Hispanic and Native groups having similar, lower odds (ie having less stressful surgeries) compared to the other groups (**Supplemental Digital Content 6**, http://links.lww.com/JACS/A276).

Increased odds of presenting with preoperative acute serious conditions, urgent cases, and emergent cases among patients from minority groups

Patients identified as Asian (adjusted odds ratio [aOR] 1.22, CI 1.17 to 1.28, $p < 0.001$, Table 1), Black (aOR 1.60, CI 1.57 to 1.64, $p < 0.001$), Hispanic (aOR 1.62, CI 1.58 to 1.66, $p < 0.001$), or Native (aOR 1.74, CI 1.62 to 1.87, $p < 0.001$) had higher odds of presenting with PASC compared to White patients. Asian, Black, and Hispanic groups also had higher odds of urgent (reference group = elective) cases vs White groups (Table 1), but not the Native group. In contrast, all 4 minority groups had higher odds of emergent cases vs the White group.

Sensitivity analyses were performed to assess the association of emergent surgeries among the race/ethnicity groups using elective vs elective/urgent cases as the reference group. The difference between the White group and the Black and Hispanic minority groups was smaller when the reference group included urgent cases, meaning that when urgent cases were included in the reference group, the odds ratio associated with Black and Hispanic groups vs White for undergoing emergent cases becomes smaller than when compared to a purely elective reference group (Fig. 1A), and remained significant after adjusting for PASC (Fig. 1B).

Worse desirability of outcome ranking outcomes among Black and Native groups and better desirability of outcome ranking outcomes among Hispanic group after adjusting for case status

We used nested models to assess the aOR of having worse (eg higher) DOOR outcomes using patients identifying as White as the reference group. Patients identified as Black and Native had worse DOOR outcomes controlling for frailty and operative stress (Table 2, M1 model), and this association endured even after controlling for case status (Table 2, M2 model) and PASC (Table 2, M3 model). The Hispanic group initially demonstrated worse DOOR outcomes (aOR 1.11, CI 1.10 to 1.13, $p < 0.001$, Table 2 M1), but this association reversed after adjusting for case status (aOR 0.96, CI 0.95 to 0.97, $p < 0.001$, Table 2 M2) and PASC (aOR 0.94, CI 0.93 to 0.96, $p < 0.001$, Table 2 M3) where they demonstrated statistically significant better outcomes compared to White groups. The Asian group demonstrated consistently better DOOR outcomes across all models (Table 2, M1 to M3).

Sensitivity analyses assessed the association of worse DOOR outcomes among the race/ethnicity groups for the Table 2 M2 and M3 models using elective vs elective/urgent cases as the reference group. Patients identified as Black had worse DOOR outcomes when the reference group was elective/urgent vs elective (Fig. 2A) and also after adjusting for PASC (Fig. 2B) compared to the White group. In contrast, the Hispanic group had higher odds of worse DOOR outcomes when elective/urgent was the reference group. However, the association reversed directionality when elective only was

Table 1. Preoperative Acute Serious Condition and Subgroups of Urgent and Emergent Case Status (Reference Group = Elective) Adjusted Odds Ratios

	PASC			Subgroup Urgent		
Predictor	**aOR**	**CI**	**p Value**	**aOR**	**CI**	**p Value**
RAI (ref = normal 21–29)						
Robust (≤20)	0.85	0.83–0.87	<0.001*	0.87	0.86–0.88	<0.001*
Frail (30–39)	2.56	2.50–2.62	<0.001*	2.99	2.95–3.02	<0.001*
Very frail (≥40)	6.52	6.25–6.80	<0.001*	6.29	6.07–6.51	<0.001*
Race/ethnicity (Ref = White, non-Hispanic)						
Asian, non-Hispanic	1.22	1.17–1.28	<0.001*	1.08	1.06–1.11	<0.001*
Black, non-Hispanic	1.60	1.57–1.64	<0.001*	1.40	1.39–1.42	<0.001*
Hispanic, any race	1.62	1.58–1.66	<0.001*	2.21	2.18–2.23	<0.001*
Native, non-Hispanic	1.74	1.62–1.87	<0.001*	1.04	1.00–1.09	0.059
PASC				10.17	9.92–10.42	<0.001*
				Emergent		
RAI (Ref = Normal 21–29)						
Robust (≤20)				1.36	1.34–1.37	<0.001*
Frail (30–39)				2.44	2.40–2.49	<0.001*
Very frail (≥40)				4.64	4.42–4.87	<0.001*
Race/ethnicity (Ref = White, non-Hispanic)						
Asian, non-Hispanic				1.91	1.87–1.95	<0.001*
Black, non-Hispanic				1.15	1.13–1.17	<0.001*
Hispanic, any race				2.18	2.15–2.21	<0.001*
Native, non-Hispanic				1.17	1.11–1.23	<0.001*
PASC				27.23	26.59–27.88	<0.001*

*Statistically significant.
aOR, adjusted odds ratio; PASC, preoperative acute serious condition; RAI, risk analysis index; ref, reference

the reference group, meaning that Hispanic patients had better (more desirable) DOOR outcomes compared to White patients (Fig. 2A). Patients identified as Asian and Native had similar odds of DOOR outcomes between the 2 reference groups.

Hispanic group has worse desirability of outcome ranking outcomes for elective cases, but better desirability of outcome ranking outcomes for urgent and emergent subgroups

The Hispanic group had increased odds of worse DOOR outcomes for elective cases (aORs range 1.05 to 1.06, **Supplemental Digital Content 7, M1 and M2**, http://links.lww.com/JACS/A276) but decreased odds (more desirable outcomes) for urgent and emergent cases (aORs range 0.84 to 0.89) compared to White groups. Patients identified as Black (aORs range 1.23 to 1.31, **Supplemental Digital Content 7, M1**, http://links.lww.com/JACS/A276) and Native (aORs range 1.11 to 1.31) vs patients identified as White had worse DOOR outcomes for each subgroup, although patients identified as Asian had better DOOR outcomes for all 3 case status subgroups.

DISCUSSION

This study demonstrates the difference between independently adjusting for elective, urgent, and emergency surgeries vs using a combined elective and urgent reference group by using a surgical DOOR. Although each minority group had improved outcomes when urgent was included as a separate category, the effects were most profound for patients identified as Hispanic and Black. In fact, patients identified as Hispanic had improved vs worse outcomes compared to patients identified as White after adjusting for case status when the reference group was elective vs combined elective and urgent (Fig. 2). Consistent with previous literature,[5-7] we have shown that best practice should identify urgent cases. Importantly, urgent cases should not be combined with elective cases and used as a reference group. Because minority populations have higher rates of both emergent and urgent surgeries,[3,11-13] this approach to risk adjustment may disproportionately impact safety-net

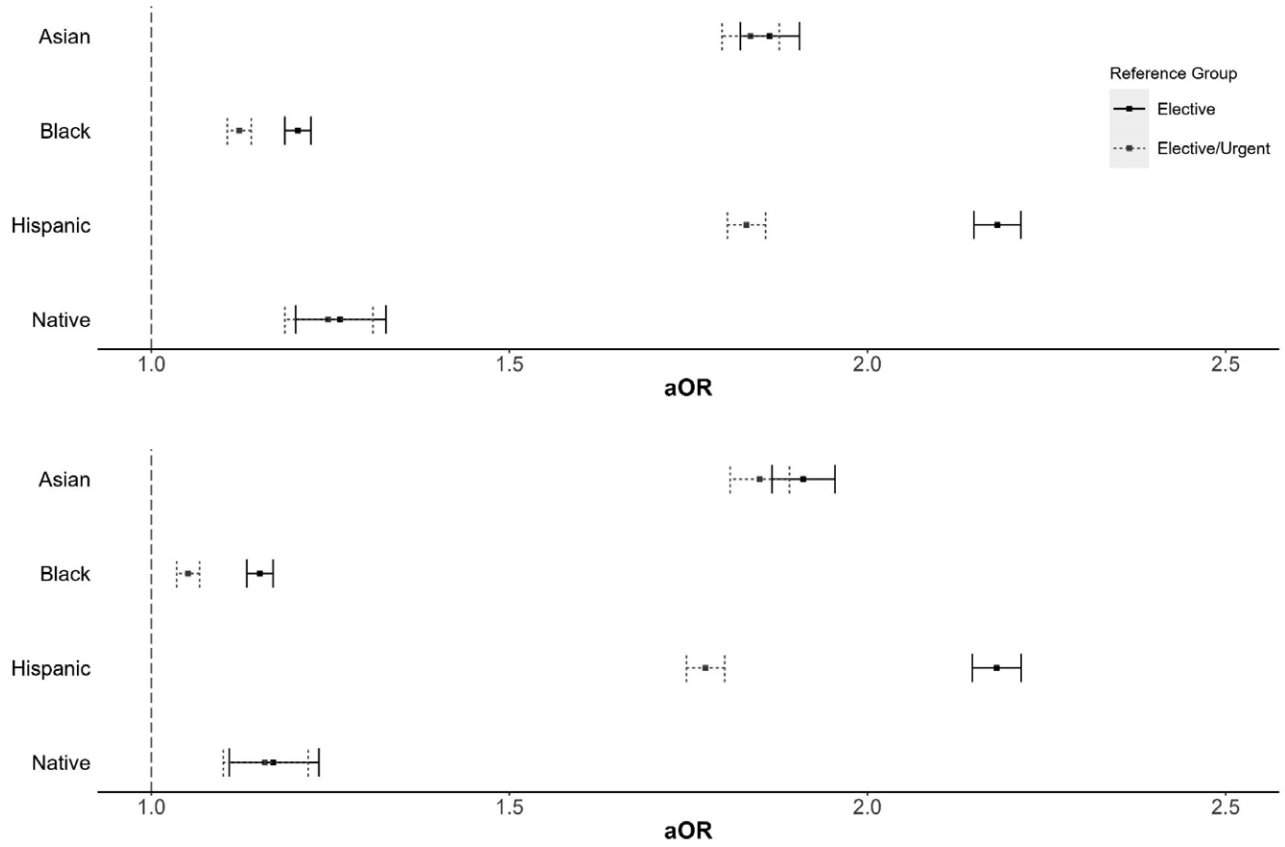

Figure 1. Racial/ethnic group differences in adjusted odds ratios of undergoing emergent surgeries using elective vs elective/urgent as the reference group sensitivity analyses. Forest plot of adjusted odds ratios (aOR) of undergoing emergent surgeries comparing different race/ethnicity groups compared to the White group. Emergent cases (Table 1, reference group = elective) were adjusted for frailty/risk analysis index (A) and risk analysis index and presenting with preoperative acute serious conditions (B) compared to using elective/urgent as the reference group. Each race/ethnicity group had odds of emergent surgery closer to the White group when using elective/urgent as the reference group. Black and Hispanic groups had the largest differences between elective and elective/urgent reference groups.

hospitals, which serve higher proportions of patients identifying as Hispanic and Black.[13] In addition, patients who have Medicare, Medicaid, or are uninsured also have higher rates of both emergent and urgent surgeries compared to patients with private insurance.[1,12] Urgent cases are not included in Pay for Performance risk adjustment. Failure to adjust for urgent surgeries may reinforce the existing disadvantages faced by minority-serving hospitals.[43-48] These factors potentially shift reimbursements from higher to lower burden safety-net hospitals, further limiting funds to treat disadvantaged populations. Administrative claims data should separate surgeries into elective, urgent, and emergent procedures to improve risk adjustment.

This study also highlights the importance of differentiating acute preoperative conditions from postoperative complications. Presenting with PASC was associated with an aOR of 13.4 for higher/worse DOOR outcomes. Administrative data use ICD-9/10 codes, which vary across healthcare systems and often cannot distinguish between preoperative risk factors (eg preoperative pneumonia or acute renal failure) from postoperative complications. NSQIP variables clearly differentiate preoperative risk factors from postoperative complications.[25,26] Administrative datasets and outcome studies should reliably differentiate preoperative acute conditions from postoperative complications to improve risk adjustment.

We developed a new surgical outcome ranking system for NSQIP PUF data using the DOOR methodology initially developed for antibiotic trials.[22-24] DOOR has a potential sensitivity advantage over other surgical outcome measures because ranks are more sensitive to effects than binary variables.[20,21] The increased aOR of higher/worse DOOR outcomes associated with increasing RAI/frailty scores, and higher levels of surgical-induced physiologic stress further validates the usefulness of DOOR. However, the use of ranked composites, such as DOOR, has several

Table 2. Adjusted Odds Ratios for NSQIP Surgical DOOR Worse Outcomes Using Nested Models (M1 to M3)

Predictor	DOOR M1			DOOR M2			DOOR M3		
	aOR	CI	p Value	aOR	CI	p Value	aOR	CI	p Value
RAI (ref = Normal 21-29)									
Robust (≤20)	0.71	0.70–0.72	<0.001*	0.71	0.70–0.71	<0.001*	0.70	0.70–0.71	<0.001*
Frail (30–39)	2.11	2.09–2.14	<0.001*	1.76	1.74–1.78	<0.001*	1.76	1.74–1.78	<0.001*
Very frail (≥40)	4.77	4.64–4.90	<0.001*	3.41	3.32–3.51	<0.001*	2.98	2.90–3.07	<0.001*
Expanded OSS (ref = OSS3)									
OSS1-2†	0.71	0.71–0.72	<0.001*	0.67	0.67–0.68	<0.001*	0.62	0.62–0.63	<0.001*
OSS4	2.42	2.40–2.44	<0.001*	2.46	2.44–2.48	<0.001*	2.12	2.10–2.14	<0.001*
OSS5	3.81	3.73–3.88	<0.001*	4.64	4.55–4.73	<0.001*	4.36	4.27–4.44	<0.001*
Race/ethnicity (ref = White, non-Hispanic)									
Asian, non-Hispanic	0.96	0.94–0.97	<0.001*	0.89	0.88–0.91	<0.001*	0.90	0.89–0.92	<0.001*
Black, non-Hispanic	1.34	1.33–1.35	<0.001*	1.28	1.27–1.29	<0.001*	1.23	1.22–1.25	<0.001*
Hispanic, any race	1.11	1.10–1.13	<0.001*	0.96	0.95–0.97	<0.001*	0.94	0.93–0.96	<0.001*
Native, non-Hispanic	1.17	1.13–1.21	<0.001*	1.15	1.10–1.19	<0.001*	1.07	1.03–1.12	<0.001*
Case status (ref = elective)									
Urgent				2.25	2.23–2.27	<0.001*	1.81	1.80–1.83	<0.001*
Emergent				3.21	3.18–3.24	<0.001*	1.95	1.93–1.97	<0.001*
PASC							13.40	13.20–13.60	<0.001*

*Statistically significant.
†OSS1 and OSS2 (very low and low stress surgeries) were combined due to small sample size of OSS1 procedures.
aOR, adjusted odds ratio; DOOR, desirability of outcome ranking; OSS, operative stress score; OSS3, moderate stress; OSS4, high stress; OSS5, very high stress; PASC, preoperative acute serious conditions; RAI, risk analysis index; Ref, reference.

concerns. First, selection or weighting of variables used in the composite can produce bias favoring a tested drug or treatment.[49] Second, outcomes involving clinical judgment, such as readmissions, are often unblinded to treatment allocation.[50] Our NSQIP surgical DOOR reduces these biases through its development in retrospective cohort data, without considering a specific drug or treatment. Third, the current surgical DOOR is limited to surgeons' consensus regarding perceived patient preference. Using consensus methods for patient groups may strengthen DOOR's applicability to patient satisfaction.

Moreover, these results demonstrate a need to understand complex health disparity issues and the advantages of using an ordinal ranking system such as DOOR. Consistently worse outcomes for patients identified as Black compared to White have been observed in various medical domains: 90-day postdischarge mortality for spinal surgery,[51] time to diagnose cervical spinal stenosis,[52] and patient ratings of access to colorectal and lung cancer care.[53] Several explanations have been offered: differences in provided care, such as minimally invasive surgery access,[54] patient mistrust of the healthcare system,[55] systemic racism,[56] and differences in presentation acuity.[57] However, researchers suggesting health disparities are driven by differences in care or systemic racism should adjust for presentation acuity, because these variables are clinically relevant, but heavily driven by race disparities and social determinants of health. For example, a study of esophagectomies and gastrectomies found that patients identifying as Black were more likely to receive open esophagectomies vs minimally invasive procedures.[54] The authors interpreted this as a care difference but failed to adjust for presentation acuity, which could have driven their results. In contrast, a study on colorectal surgery outcomes demonstrated that patients who were on Medicaid or were uninsured had higher rates of undergoing open vs laparoscopic procedures compared to patients with private insurance, but similar odds of open procedures after adjusting for PASC and urgent/emergent case status.[12] We encourage researchers to use PASC[2,4,12] and elective/urgent/emergent case status categories for surgical outcome studies or a measure of presentation acuity specific to a family of procedures, such as using CPT codes to differentiate appendicitis from appendiceal rupture or abscess. Importantly, this study demonstrated that, in addition to improving quality of care during the hospitalization, improving surgical outcomes, especially in ethnic/minority populations, must also focus on decreasing presentation acuity. This could possibly be achieved by increasing access to primary and preventive care to potentially avoid urgent or emergent cases.

Patients identified as Hispanic had worse/higher DOOR outcomes than patients identified as White for

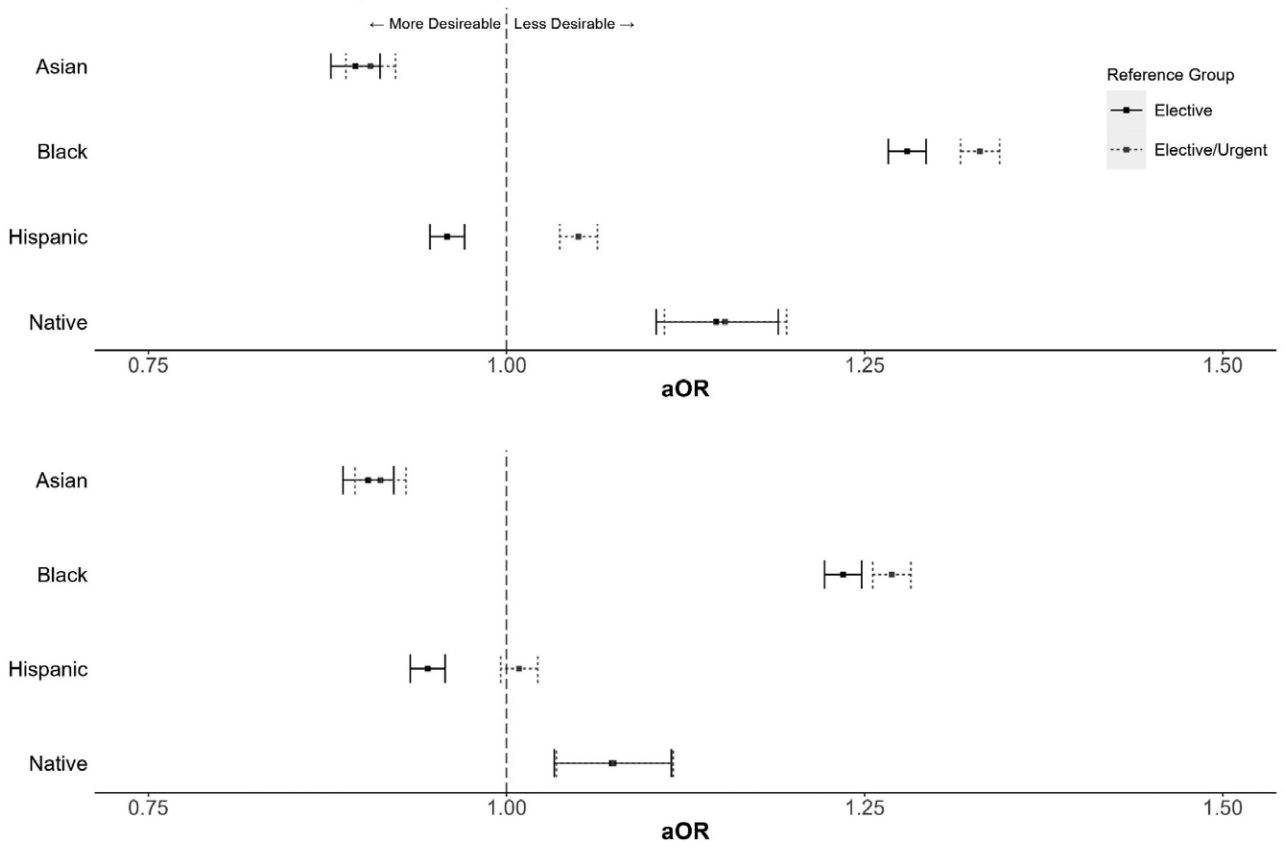

Figure 2. Racial/ethnic group differences in adjusted odds ratios of DOOR outcomes using elective vs elective/urgent as the reference group sensitivity analyses. Forest plot of adjusted odds ratios (aOR) of less desirable/worse DOOR outcomes comparing different race/ethnicity groups vs White group. DOOR outcomes (Table 2, M2 and M3, reference group = Elective) were adjusted for frailty/risk analysis index, expanded operative stress score, and case status (M2; A) and risk analysis index, expanded operative stress score, case status, and presenting with preoperative acute serious conditions (M3; B) compared to using elective/urgent as the reference group. Patients identified as Black had improved DOOR outcomes when elective only was used as the reference group compared to the combined elective/urgent reference group. Patients identified as Hispanic had worse DOOR outcomes compared to patients identified as White when using the elective/urgent reference group and the association reversed directionality, meaning that Hispanic patients had better DOOR outcomes compared to White patients when elective only cases were used as the reference group. Patients identified as Black and Hispanic demonstrated the largest differences in DOOR outcomes between elective and elective/urgent reference groups. DOOR, desirability of outcome ranking.

elective surgeries but better outcomes for urgent and emergent procedures. One possible explanation for these effects may be the high levels of social support within Hispanic family systems, which declines as socioeconomic status increases.[58] Patients with higher socioeconomic status may be more likely to receive elective surgeries but may lack the familial and social support experienced by patients with lower socioeconomic status and identifying as Hispanic.

Patients identified as White compared to the racial/ethnic minority groups were older, frailer and underwent more physiologically stressful surgeries. In contrast, the White group also had lower rates and odds of presenting with PASC and undergoing urgent and emergent cases. Interestingly, despite presenting with higher acuity, patients identified as Asian had better outcomes vs White patients for all case status subgroups, which mirrors findings that Asian patients have fewer emergency surgery complications.[59]

Our NSQIP surgical DOOR is widely useable, because it only requires variables from the NSQIP PUF. Additionally, we have provided a roadmap for others to develop DOOR systems in Procedure Targeted NSQIP PUFs.[60] Future uses could include procedure-specific variables, such as positive margins for colectomies or need to amputate for lower extremity vascular procedures, which can either be developed into new DOOR systems or included in the current DOOR. A limitation of NSQIP PUF is the lack of individual (eg insurance type) or contextual (eg Area Deprivation Index) socioeconmic status variables. Area Deprivation Index uses patients' home

addresses to differentiate more affluent vs highly deprived neighborhoods.[61] However, the racial and ethnic associations reported here persisted even after adjusting for insurance type and the Area Deprivation Index in an augmented NSQIP dataset from 3 academic healthcare systems.[62] The ACS has access to individual patient addresses and has published one abstract using NSQIP PUF data merged with the Area Deprivation Index.[63] However, the Area Deprivation Index or other address-linked variables are not present in the NSQIP PUF. We urge the ACS to add the Area Deprivation Index and other variables associated with socioeconomic status to the NSQIP PUF.

Limitations

This retrospective, longitudinal cohort study established associations with limited causal inference. Alternate, unmeasured clinical and social factors could confound our findings because NSQIP does not include variables defining patient insurance type, socioeconomic status, or education level. NSQIP provides a systematic sample of a broad range of surgeries but does not include all procedures for a healthcare system. High complication and mortality rates may occur in cases performed for palliation, rather than for the purpose of extending life, but these data do not clearly define procedures that were performed specifically for palliation. NSQIP surgical DOOR was developed by consensus rating of NSQIP complications based on surgeons' perceptions of their patients' preferences; the patient centeredness of these ratings might be improved by direct inclusion of patients' explicit perspectives, especially patients from diverse racial, ethnic, and socioeconomic backgrounds.

One assumption of proportional odds regression is that predictor effects remain similar across adjacent ranks. Clinical variables, such as PASC, have larger effects on more severe complications (higher DOOR outcomes) than for lower ranks, such as a urinary tract infection (DOOR = 2). However, these models still provide useful averages for effects across outcome severity. DOOR provides a more granular model of quality of care, rather than a means of calculating individual patient risk, because models predicting between 6 vs 2 possible categories have reduced discriminative power.

CONCLUSIONS

NSQIP surgical DOOR is a new method to increase the granularity and sensitivity of outcomes assessment and reveals complex interplay between race/ethnicity and presentation acuity. Our results suggest that presentation acuity, ie PASC and granular case status, is a key determinant of worse outcomes in patients identified as Black, Hispanic, and Native. Urgent cases had worse outcomes in all racial/ethnic groups vs elective cases. Combining elective and urgent cases in risk adjustment may disproportionately penalize safety-net hospitals that serve higher proportions of minority populations. Best practices should include risk adjusting for urgent and emergent cases. NSQIP surgical DOOR can be used to improve detection of health disparities and serves as a roadmap for developing other ordinal outcomes measures. Improving surgical outcomes should focus on decreasing PASC and urgent and emergent surgery rates, possibly by improving access to care, especially for minority populations.

Author Contributions

Dr Shireman, Mr Jacobs, and Ms Manuel had full access to the deidentified NSQIP data and take responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Shireman, Schmidt, Jacobs, and Hall
Acquisition, analysis, or interpretation of data: All authors
Drafting of the manuscript: Jacobs and Shireman
Critical revision of the manuscript for important intellectual content: All authors
Statistical analysis: Jacobs and Wang
Obtained funding: Shireman, Hall, Stitzenberg, Kao, Schmidt, and Wang
Administrative, technical, or material support: All authors
Supervision: Shireman

Acknowledgment: We thank Kristi Hill-Herrera, BSN, RN, MS, Director, Registry & Quality Assessment/Performance Improvement at University Health, for generously sharing her valuable NSQIP knowledge. The data used in this manuscript are restricted by the American College of Surgeons for use only by participating NSQIP sites. Researchers from qualified NSQIP sites can request the data directly from the American College of Surgeons.

REFERENCES

1. Simon RC, Kim J, Schmidt S, et al. Association of insurance type with inpatient surgery 30-day complications and costs. J Surg Res 2023;282:22–33.
2. Yan Q, Kim J, Hall DE, et al. Sex-related differences in acuity and postoperative complications, mortality and failure to rescue. J Surg Res 2023;282:34–46.
3. Jacobs MA, Kim J, Tetley JC, et al. Cost of failure to achieve textbook outcomes: association of insurance type with outcomes and cumulative cost for inpatient surgery. J Am Coll Surg 2023;236:352–364.

4. Yan Q, Kim J, Hall DE, et al. Association of frailty and the expanded operative stress score with preoperative acute serious conditions, complications and mortality in males compared to females: a retrospective observational study. Ann Surg 2023;277:e294–e304.
5. Mullen MG, Michaels AD, Mehaffey JH, et al. Risk associated with complications and mortality after urgent surgery vs elective and emergency surgery: implications for defining "quality" and reporting outcomes for urgent surgery. JAMA Surg 2017;152:768–774.
6. Ziegler MA, Bauman JC, Welsh RJ, Wasvary HJ. Can the American College of Surgeons National Surgical Quality Improvement Program risk calculator predict outcomes for urgent colectomies? Am Surg 2022;88:65–69.
7. Hajirawala L, Leonardi C, Orangio G, et al. Urgent inpatient colectomy carries a higher morbidity and mortality than elective surgery. J Surg Res 2021;268:394–404.
8. Havens JM, Peetz AB, Do WS, et al. The excess morbidity and mortality of emergency general surgery. J Trauma Acute Care Surg 2015;78:306–311.
9. Schwartz DA, Hui X, Schneider EB, et al. Worse outcomes among uninsured general surgery patients: does the need for an emergency operation explain these disparities? Surgery 2014;156:345–351.
10. Smith M, Hussain A, Xiao J, et al. The importance of improving the quality of emergency surgery for a regional quality collaborative. Ann Surg 2013;257:596–602.
11. Jacobs MA, Tetley JC, Kim J, et al. Association of cumulative colorectal surgery hospital costs, readmissions, and emergency department/observation stays with insurance type. J Gastrointest Surg 2023;27:965–979.
12. Tetley JC, Jacobs MA, Kim J, et al. Association of insurance type with colorectal surgery outcomes and costs at a safety-net hospital. Ann Surg Open 2022;3:e215.
13. Hoehn RS, Wima K, Vestal MA, et al. Effect of hospital safety-net burden on cost and outcomes after surgery. JAMA Surg 2016;151:120–128.
14. Ford CL, Harawa NT. A new conceptualization of ethnicity for social epidemiologic and health equity research. Soc Sci Med 2010;71:251–258.
15. Shwartz M, Restuccia JD, Rosen AK. Composite measures of health care provider performance: a description of approaches. Milbank Q 2015;93:788–825.
16. Görgec B, Benedetti Cacciaguerra A, Lanari J, et al. Assessment of textbook outcome in laparoscopic and open liver surgery. JAMA Surg 2021;156:e212064.
17. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. Ann Thorac Surg 2001;72:2155–2168.
18. McCoy CE. Understanding the use of composite endpoints in clinical trials. West J Emerg Med 2018;19:631–634.
19. van Roessel S, Mackay TM, van Dieren S, et al; Dutch Pancreatic Cancer Group. Textbook outcome: nationwide analysis of a novel quality measure in pancreatic surgery. Ann Surg 2020;271:155–162.
20. Dickey AS, Krafty RT, Pedersen NP. Ordinal regression increases statistical power to predict epilepsy surgical outcomes. Epilepsia Open 2022;7:344–349.
21. Whitehead J. Sample size calculations for ordered categorical data. Stat Med 1993;12:2257–2271.
22. Chammas M, Pust GD, Hatton G, et al; EAST Appendicitis Study Group. Outcomes of restricted vs liberal post-operative antibiotic use in patients undergoing appendectomy: a DOOR/RADAR post hoc analysis of the EAST Appendicitis MUSTANG Study. Surg Infect (Larchmt) 2022;23:489–494.
23. Barlow A, Heil EL, Claeys KC. Using an ordinal approach to compare outcomes between vancomycin vs ceftaroline or daptomycin in MRSA bloodstream infection. Infect Dis Ther 2021;10:605–612.
24. Evans SR, Rubin D, Follmann D, et al. Desirability of outcome ranking (DOOR) and response adjusted for duration of antibiotic risk (RADAR). Clin Infect Dis 2015;61:800–806.
25. Lawson EH, Zingmond DS, Hall BL, et al. Comparison between clinical registry and Medicare claims data on the classification of hospital quality of surgical care. Ann Surg 2015;261:290–296.
26. Lawson EH, Louie R, Zingmond DS, et al. A comparison of clinical registry vs administrative claims data for reporting of 30-day surgical complications. Ann Surg 2012;256:973–981.
27. Jollis JG, Ancukiewicz M, DeLong ER, et al. Discordance of databases designed for claims payment vs clinical information systems. Implications for outcomes research. Ann Intern Med 1993;119:844–850.
28. Portuondo JI, Harris AHS, Massarweh NN. Using administrative codes to measure health care quality. JAMA 2022;328:825–826.
29. Kosar CM, Loomer L, Thomas KS, et al. Association of diagnosis coding with differences in risk-adjusted short-term mortality between critical access and non-critical access hospitals. JAMA 2020;324:481–487.
30. Ghaferi AA, Schwartz TA, Pawlik TM. STROBE reporting guidelines for observational studies. JAMA Surg 2021;156:577–578.
31. Arya S, Varley P, Youk A, et al. Recalibration and external validation of the risk analysis index: a surgical frailty assessment tool. Ann Surg 2020;272:996–1005.
32. 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;152:175–182.
33. Rothenberg KA, George EL, Trickey AW, et al. Assessment of the risk analysis index for prediction of mortality, major complications, and length of stay in patients who underwent vascular surgery. Ann Vasc Surg 2020;66:442–453.
34. George EL, Massarweh NN, Youk A, et al. Comparing veterans affairs and private sector perioperative outcomes after noncardiac surgery. JAMA Surg 2022;157:231–239.
35. Reitz KM, Varley PR, Liang NL, et al. The correlation between case total work relative value unit, operative stress, and patient frailty: retrospective cohort study. Ann Surg 2021;274:637–645.
36. Yan Q, Kim J, Hall DE, et al. Association of frailty and the expanded operative stress score with preoperative acute serious conditions, complications and mortality in males compared to females: a retrospective observational study. Ann Surg 2023;277:e294–e304.
37. Shinall MC Jr, Youk A, Massarweh NN, et al. Association of preoperative frailty and operative stress with mortality after elective vs emergency surgery. JAMA Netw Open 2020;3:e2010358.

38. American College of Surgeons. User Guide for the 2014 ACS NSQIP Participant Use Data File (PUF). 2015. Available at: https://www.facs.org/media/hl0hfrxz/nsqip_puf_user-guide_2014.pdf, Accessed January 26, 2023.
39. Flanagin A, Frey T, Christiansen SL; AMA Manual of Style Committee. Updated guidance on the reporting of race and ethnicity in medical and science journals. JAMA 2021;326:621–627.
40. National Museum of the American Indian. The impact of words and tips for using appropriate terminology: am I using the right word? Available at: https://americanindian.si.edu/nk360/informational/impact-words-tips. Accessed October 12, 2022.
41. Powell C. The Delphi technique: myths and realities. J Adv Nurs 2003;41:376–382.
42. Schwarze ML, Barnato AE, Rathouz PJ, et al. Development of a list of high-risk operations for patients 65 years and older. JAMA Surg 2015;150:325–331.
43. Zogg CK, Thumma JR, Ryan AM, Dimick JB. Medicare's hospital acquired condition reduction program disproportionately affects minority-serving hospitals: variation by race, socioeconomic status, and disproportionate share hospital payment receipt. Ann Surg 2020;271:985–993.
44. Favini N, Hockenberry JM, Gilman M, et al. Comparative trends in payment adjustments between safety-net and other hospitals since the introduction of the hospital readmission reduction program and value-based purchasing. JAMA 2017;317:1578–1580.
45. Doran T, Maurer KA, Ryan AM. Impact of provider incentives on quality and value of health care. Annu Rev Public Health 2017;38:449–465.
46. Chee TT, Ryan AM, Wasfy JH, Borden WB. Current state of value-based purchasing programs. Circulation 2016;133:2197–2205.
47. Ryan AM, Burgess JF Jr, Pesko MF, et al. The early effects of Medicare's mandatory hospital pay-for-performance program. Health Serv Res 2015;50:81–97.
48. Joynt KE, Jha AK. Characteristics of hospitals receiving penalties under the hospital readmissions reduction program. JAMA 2013;309:342–343.
49. Montori VM, Permanyer-Miralda G, Ferreira-González I, et al. Validity of composite end points in clinical trials. BMJ 2005;330:594–596.
50. Cordoba G, Schwartz L, Woloshin S, et al. Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. BMJ 2010;341:c3920–c3920.
51. Engler ID, Vasavada KD, Vanneman ME, et al. Do community-level disadvantages account for racial disparities in the safety of spine surgery? A large database study based on Medicare claims. Clin Orthop Relat Res 2023;481:268–278.
52. Elsamadicy AA, Adogwa O, Fialkoff J, et al. Race as an independent predictor of temporal delay in time to diagnosis and treatment in patients with cervical stenosis: a study of 133 patients with anterior cervical discectomy and fusion. World Neurosurg 2016;96:107–110.
53. Wardrop RC, Cass AL, Quinn SA, et al. Racial and ethnic differences in patient ratings of colorectal and non-small-cell lung cancer care: a SEER-CAHPS study. Cancer Causes Control 2022;33:1125–1133.
54. Haider SF, Ma S, Xia W, et al. Racial disparities in minimally invasive esophagectomy and gastrectomy for upper GI malignancies. Surg Endosc 2022;36:9355–9363.
55. Eiser AR, Ellis G. Viewpoint: cultural competence and the African American experience with health care: the case for specific content in cross-cultural education. Acad Med 2007;82:176–183.
56. Salles A, Arora VM, Mitchell KA. Everyone must address anti-Black racism in health care: steps for non-Black health care professionals to take. JAMA 2021;326:601–602.
57. Winestone LE, Getz KD, Miller TP, et al. The role of acuity of illness at presentation in early mortality in black children with acute myeloid leukemia. Am J Hematol 2017;92:141–148.
58. Almeida J, Molnar BE, Kawachi I, Subramanian SV. Ethnicity and nativity status as determinants of perceived social support: testing the concept of familism. Soc Sci Med 2009;68:1852–1858.
59. Causey MW, McVay D, Hatch Q, et al. The impact of race on outcomes following emergency surgery: an American College of Surgeons National Surgical Quality Improvement Program assessment. Am J Surg 2013;206:172–179.
60. Mathes T, Pieper D, Morche J, et al. Pay for performance for hospitals. Cochrane Database Syst Rev 2019;7:CD011156.
61. Kind AJH, Buckingham WR. Making neighborhood-disadvantage metrics accessible - the neighborhood atlas. N Engl J Med 2018;378:2456–2458.
62. Schmidt S, Kim J, Jacobs MA, et al. Independent associations of neighborhood deprivation and patient-level social determinants of health with textbook outcomes after inpatient surgery. Ann Surg Open 2023;4:e237.
63. Fischer CP, Liu Y, Cohen ME, Ko CY. Unmeasured risks: social determinants of health affect rate of surgical complications despite clinical risk adjustment. J Am Coll Surg 2021;233:S123.

Invited Commentary

Limitations of National Databases Hinder Our Ability to Assess Surgical Outcomes and Mitigate Disparity for Minority Populations

Allison N Martin, MD, MPH
Houston, TX

During the past decade, there has been an upsurge of publications examining disparate outcomes for racial and ethnic minority patients undergoing both high acuity and elective operations. Data sources vary but include multi-institutional datasets such as the American College of Surgeons' NSQIP database, all payer datasets such as the National Inpatient Sample and administrative data, such