Effect of Wound Classification on Risk-Adjustment in American College of Surgeons NSQIP

Mila H. Ju, MD, MS[1,2], **Mark E. Cohen, PhD**[1], **Karl Y. Bilimoria, MD, MS, FACS**[1,2], **Melissa S. Latus, RN, BSN**[1], **Lisa M. Scholl, RN, MSN**[1], **Bradley J. Schwab, RN, BSN**[1], **Claudia M. Byrd, RN, BSN**[1], **Clifford Y. Ko, MD, MS, MSHS, FACS**[1,3], **E. Patchen Dellinger, MD, FACS**[4], and **Bruce L. Hall, MD, PhD, MBA, FACS**[1,5]

[1]Division of Research and Optimal Patient Care, American College of Surgeons, Chicago, IL

[2]Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University, Northwestern Memorial Hospital, Chicago, IL

[3]Department of Surgery, David Geffen School of Medicine, University of California, Los Angeles; Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA

[4]Surgery Department, University of Washington School of Medicine, Seattle, WA

[5]Department of Surgery, Washington University in Saint Louis; Olin Business School and Center for Health Policy, Washington University in St. Louis; St. Louis VA Medical Center; BJC Healthcare, St. Louis, MO

Abstract

Background—Surgical wound classification has been used in risk-adjustment models. However, it can be subjective and potentially improperly bias hospital quality comparisons. The objective is to examine the effect of wound classification on hospital performance risk-adjustment models.

Study Design—Retrospective review of the 2011 ACS NSQIP database was conducted for wound classification categories: clean, clean/contaminated, contaminated, and dirty/infected. To assess the influence of wound classification on risk-adjustment, two models were developed for each outcome: one including and one excluding wound classification. For each model, hospital postoperative complications were estimated using hierarchical multivariable regression methods. Absolute changes in hospital rank, correlations of odds-ratios, and outlier status agreement between models were examined.

Correspondence address: Mila H. Ju, MD, MS, American College of Surgeons, 633 North Saint Clair Street, 22nd floor, Chicago, IL 60611, 312.202.5042, mju@facs.org, m-ju@northwestern.edu.

Results—Of the 442,149 cases performed in 315 hospitals: 53.6% were classified as clean; 34.2% clean/contaminated; 6.7% contaminated; and 5.5% dirty/infected. The surgical site infection (SSI) rate was highest in dirty/infected (8.5%) and lowest in clean (1.8%) cases. For overall SSI, the absolute change in risk-adjusted hospital performance rank between models including vs. excluding wound classification was minimal (mean 4.5 out of 315 positions). The correlations between odds ratios of the two performance models were nearly perfect (R=0.9976, P<0.0001), and outlier status agreement was excellent (Kappa=0.9508, P<0.0001). Similar findings were observed in models of subgroups of SSI and other postoperative outcomes.

Conclusions—In circumstances where alternate information is available for risk-adjustment, there appear to be minimal differences in performance models that include vs. exclude wound classification. Therefore, ACS NSQIP is critically evaluating the continued use of wound classification in hospital performance risk-adjustment models.

INTRODUCTION

Wound classification (alternately "wound class") has an interesting background. On September 19, 1958, the National Academy of Sciences National Research Council met to design a multi-institutional study that evaluated the effect of ultraviolet irradiation in preventing operative infections. The study results were first published in the Annals of Surgery in 1964 with a further report in 1968.[1,2] The study introduced wound classification as a way to describe the degree of bacterial or infection load present at the time of surgery. The reported infection rates were much higher than current day experience, with 5.1% of clean, 10.8% of clean-contaminated, 16.3% of contaminated, and 28.6% of dirty wound class cases developing infections.[1] In 1970, the National Nosocomial Infectious Surveillance (NNIS) system was established, as an aggregated database for infections occurring at acute care hospitals, and this incorporated wound classification as a risk factor for surgical site infections (SSI).[3,4] In 1985, wound classification also appeared in the Center for Disease Control's (CDC) guidelines for prevention of postoperative wound infection as a "Must" element of documentation at the time of the operation.[5] However, none of these early sources were absolute "owners" of the system.

Although wound classification has been used as an important predictor of risk for SSI and other outcomes in quality improvement efforts since 1964, there are some major concerns surrounding its use. Practices of wound class documentation vary across the nation as there has been no absolute authority for ongoing development of, monitoring of, or provision of guidance for the classification system. For instance, among participants in the American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP), there are differences in opinion as to whether wound class should be assigned based on the "planned procedure" (e.g., laparoscopic cholecystectomy), or whether it should be assigned after procedure completion based on actual occurrences (e.g., laparoscopic cholecystectomy with inadvertent colotomy). In addition, certain aspects of common guidance on this topic might be off-target, such as classification of a clean, uninfected nephrectomy as clean/contaminated because the ureter is transected. The perception from the CDC is that the ACS should provide guidance on wound classification documentation and usage in risk-adjustment models (authors' personal communications with CDC). However, the ACS has

never claimed such responsibility for the wound classification system, and its definition is based on the 1999 CDC guideline for prevention of SSI.[6]

There is a potential concern that non-standardized categorization of wound class, and use of wound class in risk-adjustment models for quality improvement, could have unintended consequences in public reporting of hospital performance, where the provider(s) might not be held fully responsible for error in care. For example, elective cholecystectomies are considered as "clean/contaminated" cases because of transection of the potentially colonized biliary tract. If an inadvertent entry into the colon occurs during cholecystectomy, the case is either still coded as "clean/contaminated" (consistent with the "planned procedure" but will have a higher than expected risk of infection for that category) or as "contaminated" (consistent with the "procedure performed" and will have an appropriate risk of infection for this category). If the "contaminated" category is used in subsequent risk-adjusted quality assessment, the provider will now be allowed an infection rate typical of a contaminated case, even though it was contaminated because of a potential provider misadventure. If an SSI occurs in the postoperative period, the contaminated wound classification will confer to this surgical case a favorable risk-adjustment. The dilemma stems from the fact that the wound classification system was originally developed to stratify expectations for different cases, but the system is now being used for risk-adjustment.

The objective of this study is to examine the effect of wound classification on risk-adjustment models for hospital performance in the ACS NSQIP. Determining the criticality of this variable to risk performance model is necessary for making definitive recommendations about the use of wound class in light of the concerns raised above, and, in turn, for offering definitive guidance about coding of wound classification in general.

METHODS

Data Source and Patient Selection

This work involved retrospective review of the ACS NSQIP database from January to December 2011. Data acquisition and validation regarding ACS NSQIP have been previously described.[7,8] As the first clinically audited and validated multi-specialty surgical quality improvement program in the nation for hospitals in the private sector, ACS NSQIP provides participating hospitals with comparative risk-adjusted outcome reports that can be used to identify potential targets for quality improvement. These analyses included all surgical procedures in the database performed in patients 18 years or older during this period. These analyses used preexisting, patient-deidentified data.

Preoperative, Postoperative, Clinical Outcome Variables

The trained surgical clinical reviewers (SCR's) at each participating institution collect at least 130 standard ACS NSQIP variables on every case, which include preoperative, operative, and 30-day postoperative outcome data. Preoperative characteristics include patient demographics, American Society of Anesthesiologists (ASA) classification, functional status, and other comorbidities. Wound classification is a variable that is entered into the record typically inside of the operating room. Wound class might be described as a

"preoperative" characteristic of the operation performed, but if it is revised based on operative occurrences, it is no longer strictly a "preoperative" characteristic, and might be best described as an "operative" variable. It consists of four categories: clean, clean/contaminated, contaminated, and dirty/infected. According to ACS NSQIP variable definitions, clean cases are those without infection or inflammation, and in which there is no entry into the respiratory, gastrointestinal, genital, or urinary tracts. In clean/contaminated cases, these tracts are entered in a controlled condition, and there is no additional infection or inflammation encountered. Contaminated cases have open accidental wounds, major breaks in sterile technique, or gross spillage from the gastrointestinal tract. Dirty/infected cases include those with old traumatic wounds with devitalized tissues, and cases with preexisting infection or preexisting perforated viscera.

The main outcome of interest in the risk adjustment models for this work was any SSI, which includes subgroups of superficial incisional, deep incisional, and organ space infections. Superficial incisional SSI is infection that is limited to the skin or subcutaneous tissue of the surgical incision. Deep incisional SSI is infection that has spread to deep soft tissues such as fascial and muscle layers. SSI that has both superficial and deep components is documented only as deep incisional SSI. Organ space SSI is "deeper" infection that involves any part of the anatomy that was opened or manipulated during the operation, and for instance would include an abscess that develops at the surgical site in a delayed fashion. Importantly, currently in the ACS NSQIP, if patients show evidence of infection that was present at the time surgery was initiated, they are not eligible to have these complications logged subsequently (if recognized postoperatively). This is referred to as a "Present At Time Of Surgery" ("PATOS") modification), and is similar in concept to "Present on Admission" ("POA") modifiers. Furthermore, if patients have an incomplete surgical wound closure, they are only eligible for an infection to be logged at a level as deep as or deeper than the level of closure. For instance, a patient who has skin left open at the close of surgery cannot be logged as having a superficial skin infection subsequently. Both the "PATOS" and the wound closure modifications differentiate ACS NSQIP from most other evaluation programs.[9] Other 30-day postoperative outcomes were also modeled, including: sepsis or septic shock, wound disruption, cardiac event (myocardial infarction or arrest requiring cardiopulmonary resuscitation), venous thromboembolism (new diagnosis of vein thrombosis or pulmonary embolism), pneumonia, reintubation, prolonged ventilation (>= 48 hours), urinary tract infection, renal failure, reoperation, and mortality. Each of these is rigorously defined in the ACS NSQIP.[9]

In order to shed insight on a case type that has uncertainty regarding proper classification according to common guidance, the specific question of the infection risk of nephrectomy cases was separately examined. Transection of any mucosal lined ductal structure is considered as clean-contaminated at a minimum by predominant guidance. However, nephrectomy is often considered a clean case by practitioners because transection of a ureter typically exposes sterile urine. Nephrectomy cases were identified from all cases using American Medical Association Current Procedural Terminology (CPT) codes 50220, 50225, 50230, 50234, 50236, 50240, 50543, 50545, 50546, and 50548. Occurrences of any SSI and subtypes of SSI were examined separately for each of the four categories of wound classification for nephrectomy cases and for all other cases (excluding nephrectomy).

Statistical Analyses

To assess the effect of wound classification on risk-adjustment, two models were developed for each postoperative outcome. The first (No Wound Class) used standard ACS NSQIP risk-adjustment variables excluding wound classification, while the second (With Wound Class) used standard ACS NSQIP risk-adjustment variables and included wound classification (see Supplemental Table 1, online only, for list of fixed-effect variables significant for each model). As different procedures have different risks for developing postoperative complications, CPT linear risk for each outcome was included as a fixed-effect variable in each model. Hospital quality performance was estimated for each model of each postoperative outcome using a hierarchical modeling approach (random intercept, fixed slope) with hospital being a random factor. This is the current standard ACS NSQIP approach that accounts for potential clustering of patients by hospitals, and incorporates a reliability or "shrinkage" adjustment as well.[10] It produces the best estimate of hospital performance in terms of an odds ratio (OR, the odds of an event at a particular hospital compared to the odds of the event at a statistically estimated average hospital).[11,12] For each model, hospitals were ranked from best to worst performance based on the odds ratio point estimate. A hospital was considered a performance outlier if the OR point estimate and 95% confidence interval (CI) were entirely less than 1.0 (better than expected) or greater than 1.0 (worse than expected).

The differences in hospital quality performance assessments under the "No Wound Class" and "With Wound Class" models were assessed in five ways: 1) correlation between model odds ratios; 2) changes in outlier status; 3) change in rank (mean, median); 4) change in percentile (mean, median); and 5) change in decile (mean, median). First, correlations between the log-transformed hospital specific odds ratio point estimates between models were assessed using Pearson correlation coefficients. Log-transformation was performed to normalize the distribution of odds ratios. Second, hospital outlier status agreement between models was compared using the Cohen's weighted kappa statistic, with values of >0.75 representing excellent agreement.[13] In addition, the absolute differences in performance rank number (ranks 1–315) and percentile (1–100) between models were examined in terms of mean, median, and interquartile range (IQR) for change. Finally, absolute changes in performance deciles between models were also examined in terms of mean, median, and IQR for change. This was to give a more realistic estimate of impact on "leagues" or categories of performance, and to blunt the effect of small, potentially inconsequential changes in rank.

Model discrimination and calibration were also examined using c-statistics (1 is perfect model discrimination) and Brier score (0 is perfect model discrimination and calibration).[10] Statistical significance was set at P<0.05 based on 2-sided tests. Statistical analyses were performed using SAS 9.3 software (SAS Institute, Cary, NC).

RESULTS

The 2011 ACS NSQIP database included 442,149 cases performed in 315 hospitals. Among these cases, 237,002 (53.6%) were classified as clean, 151,213 (34.2%) were clean/contaminated, 29,495 (6.7%) were contaminated, and 24,439 (5.5%) were dirty/infected

(Figure 1). Preoperative patient characteristics were generally similar among wound classes but in some instances patients with dirty/infected wound class had higher prevalence of comorbidities (Supplemental Table 2, online only).

Overall, the incidence of any SSI (superficial incisional, deep incisional, or organ space) was 3.4%, with 1.9% as superficial incisional SSI, 0.6% as deep incisional SSI, and 1.1% as organ space SSI (as a case can have two types of SSI: superficial and organ space, or deep and organ space, but not superficial and deep)(Table 1). The rate of any SSI was highest among cases that had wound classification of dirty/infected (8.5%) and lowest in cases that were clean (1.8%) (Table 1 and Supplemental Figure 1, online only). The highest incidence of superficial incisional SSI occurred in the contaminated wound class (824, 2.8%), whereas the highest incidences of deep incisional and organ spa.ce SSI occurred in dirty/infected wound class (373, 1.5% and 1,079, 4.4% respectively). For all other 30-day postoperative outcome variables, dirty/infected cases had the highest rate of complications when compared to the other wound classes (Table 1). Clean cases had shortest length of stay (mean 2.4; median 1.0, IQR 0.0–3.0 days) while dirty/infected cases had longest length of stay (mean 8.1; median 5.0, IQR 2.0–10.0 days).

Nephrectomy cases were coded as 17.0% clean, 80.2% clean/contaminated, 1.7% contaminated, and 1.1% dirty/infected. Although the rates of infection in the *non-nephrectomy* cases were the similar to infection rates of the entire study cohort, isolated nephrectomy cases had infection rates of 1.5% in clean cases, 2.4% in clean/contaminated cases, and 2.9% in contaminated cases, which all were lower than the infection rates of the entire study cohort (1.8%, 4.8%, and 5.6% respectively). On the other hand, for "dirty-infected" nephrectomies, the infection rate (11.6%) was higher than for all other dirty-infected cases (8.5%).

Hospital Quality Comparisons

To determine the influence of wound classification on risk-adjustment for hospital quality comparisons, correlations between hospital performance odds ratios, changes in outlier declaration, and absolute changes in hospital rank, percentile, and decile status were compared between models with and without wound class (Tables 2 and 3). Correlations of the log-transformed hospital-specific odds ratio point estimates between models for incidence of any SSI and for SSI subgroups reflected near identity (R>0.990, P<0.0001). Outlier status agreement between models was also excellent for any SSI (Kappa=0.9508, P<0.0001), superficial incisional SSI (Kappa=0.9872, P<0.0001), deep incisional SSI (Kappa=0.9520, P<0.0001), and organ space SSI (Kappa=0.9007, P<0.0001). For the main modeled outcome of interest, incidence of any SSI, the mean absolute change in hospital performance ranking was 4.5 positions (out of 315), and median absolute change was 3.0 (IQR 1.0–6.0). Similarly, the mean and median absolute changes in rank (out of 315) for the subgroup superficial incisional SSI were 0.9 and 1.0 (IQR 0.0–2.0), for deep incision SSI were 4.2 and 3.0 (IQR 1.0–5.0), and for organ space SSI were 9.9 and 6.0 (IQR 2.0–14.0) (Figure 1). The absolute changes in hospital performance percentile were minimal for any SSI (mean 1.41 percentile positions; median 0.95, IQR 0.32–1.90), for superficial incisional SSI (mean 0.30 percentile positions; median 0.32, IQR 0.00–0.63), for deep incisional SSI

(mean 1.33 percentile positions; median 0.95, IQR 0.32–1.59), and for organ space SSI (median 3.13 percentile positions; median 1.90, IQR 0.63–4.44). In addition, the majority of hospitals (261 hospitals or 82.86%) did not change decile grouping in between the two models for any SSI, while the remaining 54 hospitals (17.14%) changed by only one decile. For each infection subgroup models, the majority of hospitals did not change decile: superficial incisional SSI (309, 98.10% no change in decile), deep incisional SSI (275, 87.30%), and organ space SSI (219, 69.52%) (Figure 2). In fact, only for organ space SSI did any hospital shift more than one decile, and the maximum shift was two deciles. In addition, the absolute changes in rank between models including and excluding wound class were similar for hospitals across deciles (median change of 2.0, 3.0, 2.0, and 2.0 for deciles one, two, nine, and ten, respectively, in the any SSI models).

Furthermore, results were comparable when other NSQIP-defined 30-day postoperative outcomes (sepsis or septic shock, wound disruption, cardiac event, venous thromboembolism, pneumonia, reintubation, prolonged ventilation, urinary tract infection, renal failure, reoperation, or mortality) were substituted as the modeled clinical outcome. The performance correlations were high; outlier status agreements were excellent; and absolute mean and median changes in hospital rankings, percentile, and decile were minimal (Tables 2 and 3).

Model discrimination and calibration were similar between models with and without wound classification. For example, the c-statistic for the "any SSI" model with wound classification was 0.832 while without was 0.829. The Brier scores were the same (0.0307) for both models. Very similar minimal differences between the two models were seen for the other postoperative outcomes.

DISCUSSION

Since its introduction in 1964, wound classification has been extensively studied and incorporated into many risk-adjustment models and quality improvement programs, including the ACS NSQIP. However, this study shows that hospital quality assessment for SSI and other 30-day postoperative outcome variables did not change substantially between the models that excluded or included wound classification in risk-adjustment, given the other patient and procedure information available within the ACS NSQIP. When hospitals were ranked based on their performance for different outcome variables, the median change in rank was minimal, around two rank positions, or less than a percent of the distribution. Moreover, the performance correlations between the two models were nearly perfect, and outlier status agreements were excellent for all of the postoperative outcomes examined.

Several studies have suggested that perhaps the four category wound classification system is flawed and needs to be reassessed. Some have argued that wound classification by itself has limited predictive power because of lack of individual patient factors in its definitions, and that multivariate models that include wound classification combined with patient variables are more predictive of SSI.[4,14] Nonetheless, those arguments still involved the current wound classification categories. Ferraz et al found that with clean cases alone, there still was a great variation in incidence of SSI for patients with different risk factors.[15] In the classic

ultraviolet light study from which wound classes were derived, the infection rate for radical mastectomy, a clean operation, was 18.9% while that for colectomy with anastomosis, a clean-contaminate operation, was 10%.[1] Ortega et al found that the current wound classification did not take into account that certain procedures were associated with higher SSI rate.[16] In addition, our work confirms the notion that even one specific procedure, such as nephrectomy, can be classified differently by different parties under differing circumstances, even in contradiction to prevailing guidance. Furthermore, such practices might be appropriate, since subsequent infection rates by class for this procedure differ from the class infection risks seen in the entire population, and the class rates for this procedure fall in between the class rates for the overall population.

Regarding actual wound class assignment to different cases, there is no single authoritative source of clinical guidance. The perception from CDC and perhaps from the public is that ACS is responsible for the wound classification system. However, ACS uses the wound class definitions in the 1999 CDC guideline for prevention of SSI.[6] Furthermore, the ACS has never made any official claim to be responsible for the system. The Association of Perioperative Registered Nurses (AORN) have recommendations on documentation of wound classes, but these lack details for specific surgical procedures.[17] Interestingly, the National Health Services (NHS) of the United Kingdom has published guidance that specifies the minimum wound class for different surgical groups in its classification of interventions and procedures, however, such guidance has not been offered or adopted in United States.[18] Notably, the NHS specifically describes this guidance as "minimum wound classes" for procedures, noting that wound class should be assigned after completion of the procedure, and should factor in all available information from the procedure. Wound class is not strictly determined by the intended procedure nor is it strictly determined by the nomenclature of the procedure, and should be assigned based on all information available from the conduct of the procedure. This guidance clearly fits certain purposes, namely yielding the most accurate prediction of infection risk for that case moving forward from that point in time. However, this also maximizes the potential to bias risk-adjustment by setting an expectation of a higher ultimate infection rate. In fact, some would argue that, since this approach relies on information that is not known prior to the procedure, and in fact not until the completion of the procedure, wound class established in this way cannot be considered a preoperative variable eligible for risk-adjustment purposes.

Without clear guidance in the United States, assignment of wound classification would depend on the aims of the classifier. If the classifier intends to judge the quality of the surgery performed, it should not unintentionally correct for potential errors occurring during the procedure (by risk-adjusting for the error). On the other hand, if the intent of the classifier is to accurately predict the risk of infection (i.e., to inform the patient or providers of risk from this point forward), then all information available at the time of the estimation should be factored into the estimation.

Of course, there is also a chance that the coding system's complexity could lead to outright errors of classification. In a study of general surgery and trauma cases, circulating nurses were able to classify general surgical wounds with 88% accuracy, but trauma cases with just 82% accuracy, when judged against the "all operative information" standard.[19] Against such

a gold standard, misclassification can occur when wound class is not updated after new intraoperative findings or a change in procedure, when there is poor communication between staff, and when there is simply a lack of knowledge of wound classification criteria. Devaney et al found that even after 3 months of teaching and audit, misclassification was only slightly reduced from 19% to 14%.[20]

Importantly, "misclassification" can also be due to practices in blatant disagreement with wound class definitions. For example, if using prevailing current guidance, nephrectomy should be classified as a clean/contaminated case because the urinary tract (a mucosal lined ductal structure) is entered or transected. However, surgeons might consider this wound to be clean because urine is typically sterile. In this ACS NSQIP-based study, although the majority of nephrectomies were classified as clean/contaminated (80.2%), 17.0% were classified as clean. Such practices might be appropriate, since subsequent infection rates by class for this procedure differ from the class infection risks seen in the entire population. The 80% of cases classified as clean/contaminated had a lower infection risk that other clean/contaminated cases in the surgical population, but a higher risk overall than clean cases. The 17% of nephrectomies classified clean had a lower infection rate than other clean cases, and contaminated nephrectomies had a lower rate than other contaminated cases. Yet when a nephrectomy was classified as dirty/infected, the risk was higher than the rest of dirty/infected cases. This is all evidence that current predominant guidance available for the profession does not apply well to all procedures, and that coding practices for such procedures likely vary across the profession. This sets the stage for bias in performance assessment. This situation lends credence to the stance that wound class coding should be tailored to the conduct of each individual case, but that the variable should then be avoided for the purpose of risk adjustment.

It would seem, therefore, that reliable utilization of wound class information would mandate either standardized guidance toward a particular specified aim, or avoidance of use of the information when it could be coded in a way that would create a bias that would defeat the user's aims. Currently, using wound classification (as currently practiced in the U.S.) in risk-adjustment for quality improvement (for instance in hospital performance reports in ACS NSQIP, or in metrics reported on CMS Hospital Compare or components of the Medicare Hospital Value-Based Purchasing Program) could have this kind of unintended consequence. When bowel is entered accidentally during an otherwise clean case, the wound classification might be changed to "contaminated". If a SSI then occurred in the postoperative period, the "contaminated" classification would risk-adjust this case in a favorable direction for the provider. This type of effect has been seen in prior work in ACS NSQIP examining hysterectomy complications, where it was seen that a wound class of "contaminated" for these cases was highly predictive of subsequent infection.[21] In other words, wound class was a powerful risk-adjuster. However, further examination revealed that commonly a "contaminated" wound class for these procedures reflected an unintended intraoperative occurrence such as inadvertent enterotomy. Risk-adjustment for such occurrences would inappropriately control for these misadventures, distorting performance evaluations by "blunting" the ability to reflect the occurrence of a bad event. In other words, the hospital would be given some "allowance" for the SSI event by making the patient appear sicker by virtue of a higher wound class, although (critically) that information was

not present or known prior to the start of the intervention. While our study was limited to ACS NSQIP data, hospital performance determinations for SSI measures reported on CMS Hospital Compare or incorporated into the Medicare Hospital Value-Based Purchasing Program (now or in the future) could also use wound classification in risk-adjustment, and therefore could be subject to the same distortions.

Given the potential to generate bias or unintended consequences for performance evaluations, it is natural to re-examine whether the information gained by wound classification is critical to risk-adjustment. This work demonstrates that when taking into account other procedural and patient risk factors, wound classification did not change quality performance assessments substantially. In addition, while the finding that wound class is a significant predictor of many different outcomes is an interesting one (dirty/infected cases had the highest rate of complications when compared to the other wound classes), across all the examined outcomes it could still be eliminated from risk-adjustment without substantial detriment. This means, of course, that there is redundant explanatory power within the other ACS NSQIP variables. Therefore, avoidance of this variable for risk-adjustment can eliminate potential unintended consequences, while minimally affecting performance evaluation. Some might argue that since the effect of the variable is minimal, the risk of bias must also be minimal, but this is an oversimplification that might trivialize the potential bias. In fact, organ space infections seem to be the most sensitive infection type to inclusion of wound class from our results; this could be interpreted to mean that for a variety of reasons this subgroup is most liable to the aforementioned bias.

Our study should be considered in light of several limitations. Risk-adjustment modeling and quality performance comparisons were performed at the hospital and not at the patient level. Wound classification might be important when considering patient level prediction of postoperative infection, but current quality improvement and pay-for-performance metrics are focused on hospital-level modeling. Selection bias might be present in this study because it was limited to hospitals that participated in ACS NSQIP. Participation in quality improvement programs requires considerable effort and resources from the hospitals. The awareness of ongoing monitoring and the existence of continuous quality improvement reduce complication rates. Wound classification may make a difference in risk-adjustment when complication rates are higher. In addition, wound classification could have more importance for other focused questions which were not examined.

Conclusions

This work reaches the following major conclusions:

1. Wound class does not appear to be consistently implemented for all surgeries. Different practices with regard to classification of wound class might not suit the intended purposes of users. In particular, wound class might create unwarranted bias in risk-adjustment performance evaluations and its use for such purpose should be carefully considered if not abandoned.

2. Most nephrectomies' risk of infection is in between "clean" and "clean/contaminated" cases, while those classified "dirty/infected" have even higher risk than expected for that category. The current wound class system does not optimize

analytic handling of these cases. Other surgeries involving transection of an endothelial lined tract might warrant "clean/contaminated" status, but in some circumstances if the tract/surface is believed to be sterile then "clean" status could still be appropriate.

3. Wound class has a fairly small effect on risk-adjustment of institutional performance evaluations in ACS NSQIP given the availability of redundant explanatory power present in other risk adjustment covariates. This is observed across a broad swath of models. Wound class could be eliminated from ACS NSQIP (and perhaps other quality performance programs) risk-adjustment algorithms without significant negative impact on quality assessment.

We believe that the most added value to classification of wounds would be obtained if procedures were coded based on the actual conduct of a case. Currently, there is great concern that wound coding practices across the country are unjustifiably variable. The "typical" wound class for a case should be a minimum that can be superseded by the clinical judgment of the operating surgeon or team, based on all information obtained and events taking place in the operating room. This standard is in line with individual patient prospective risk prediction, but not aligned with risk-adjustment for performance evaluation, where in any case it appears to add little power. Wound class should not be used for this risk-adjustment purpose.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

ABBREVIATIONS AND ACRONYMS

ACS NSQIP	American College of Surgeons National Surgical Quality Improvement Program
SSI	Surgical site infection
NNIS	National Nosocomial Infectious Surveillance
CDC	Center for Disease Control's
SCR's	Surgical clinical reviewers

| ASA | American Society of Anesthesiologists |
PATOS	Present At Time Of Surgery
POA	Present on Admission
CPT	Current Procedural Terminology
OR	Odds ratio
CI	Confidence interval
IQR	Interquartile range
AORN	Association of Perioperative Registered Nurses
NHS	National Health Services

References

1. Berard F, Gandon J. Postoperative wound infections: the influence of ultraviolet irradiation of the operating room and of various other factors. Ann Surg. 1964; 160:1–192. [PubMed: 14179433]

2. Hart D, Postlethwait RW, Brown IW Jr, et al. Postoperative wound infections: a further report on ultraviolet irradiation with comments on the recent (1964) national research council cooperative study report. Ann Surg. 1968; 167:728–743. [PubMed: 5646294]

3. National Nosocomial Infections Surveillance (NNIS) report, data summary from October 1986-April 1996, issued May 1996. A report from the National Nosocomial Infections Surveillance (NNIS) System. Am J Infection Control. 1996; 24:380–388.

4. Culver DH, Horan TC, Gaynes RP, et al. Surgical wound infection rates by wound class, operative procedure, and patient risk index. National Nosocomial Infections Surveillance System Am J Med. 1991; 91:152S–157S.

5. Garner JS. CDC guideline for prevention of surgical wound infections, 1985. Supersedes guideline for prevention of surgical wound infections published in 1982. (Originally published in November 1985). Revised Infection Control. 1986; 7:193–200.

6. Mangram AJ, Horan TC, Pearson ML, et al. Centers for Disease Control and Prevention (CDC) Hospital Infection Control Practices Advisory Committee. Guideline for Prevention of Surgical Site Infection, 1999. Am J Infection Control. 1999; 27:97–132. quiz 133–134; discussion 196.

7. Khuri SF, Henderson WG, Daley J, et al. Successful implementation of the Department of Veterans Affairs' National Surgical Quality Improvement Program in the private sector: the Patient Safety in Surgery study. Ann Surg. 2008; 248:329–336. [PubMed: 18650645]

8. Bilimoria KY, Cohen ME, Ingraham AM, et al. Effect of postdischarge morbidity and mortality on comparisons of hospital surgical quality. Ann Surg. 2010; 252:183–190. [PubMed: 20531000]

9. [Accessed May 5, 2013] American College of Surgeons National Surgical Quality Improvement Program (ACS NSQIP). http://site.acsnsqip.org/

10. Cohen ME, Ko CY, Bilimoria KY, et al. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. J Am Coll Surg. Apr 26.2013

11. Greenland S. Principles of multilevel modelling. Int J Epidemiol. 2000; 29:158–167. [PubMed: 10750618]

12. Moore L, Hanley JA, Turgeon AF, Lavoie A. Evaluating the performance of trauma centers: hierarchical modeling should be used. J Trauma. 2010; 69:1132–1137. [PubMed: 20404760]

13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometr. 1977; 33(1):159–174.

14. Haley RW, Culver DH, Morgan WM, et al. Identifying patients at high risk of surgical wound infection. A simple multivariate index of patient susceptibility and wound contamination. Am J Epidemiol. 1985; 121:206–215. [PubMed: 4014116]

15. Ferraz EM, Bacelar TS, Aguiar JL, et al. Wound infection rates in clean surgery: a potentially misleading risk classification. Infect Control Hosp Epidemiol. 1992; 13:457–462. [PubMed: 1517544]

16. Ortega G, Rhee DS, Papandria DJ, et al. An evaluation of surgical site infections by wound classification system using the ACS-NSQIP. J Surg Research. 2012; 174:33–38. [PubMed: 21962737]

17. Association of periOperative Registered Nurses. Recommended practices for documentation of perioperative nursing care. AORN. 2000; 71:247–250.

18. (SSISS) NHSCfHSSISS. OPCS Classification of Interventions and Procedures. Version 4.5. London, UK: TSO; 2011. Protocol for surveillance of surgical site infection.

19. Cardo DM, Falk PS, Mayhall CG. Validation of surgical wound classification in the operating room. Infect Control Hosp Epidemiol. 1993; 14:255–259. [PubMed: 8496578]

20. Devaney L, Rowell KS. Improving surgical wound classification--why it matters. AORN. 2004; 80:208–209. 212–223.

21. Gashaw Lake A, McPencow AM, Dick-Biascoechea MA, et al. Surgical site infection after hysterectomy. Am J Obstet Gynecol. Jun 13.2013
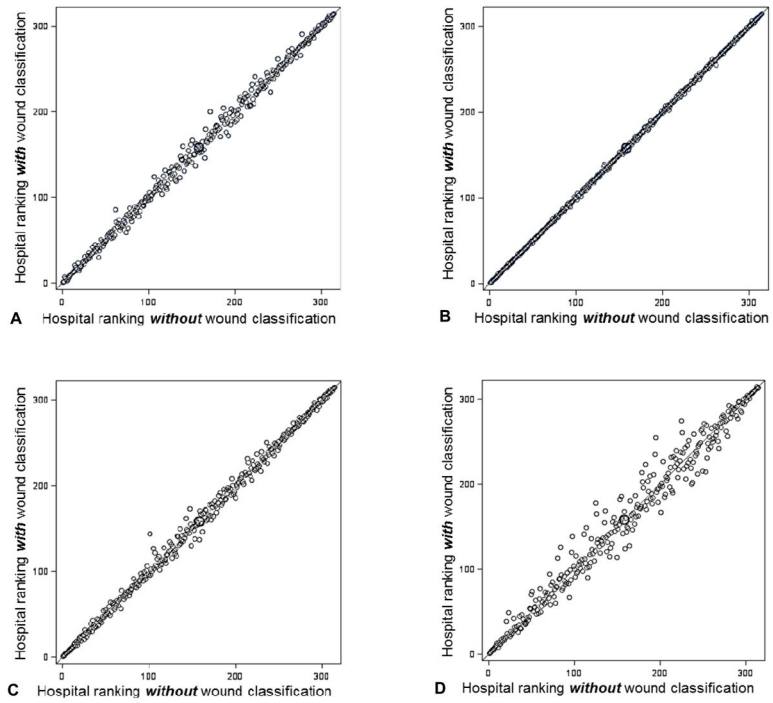
Figure 1.
Comparison in ranks for 315 hospitals in NSQIP (2011) using the model *excluding* wound classification and the model *including* wound classification for (A) any surgical site infection (SSI), (B) superficial incisional SSI, (C) deep incisional SSI, and (D) organ space SSI. Each circle represents 1 of the 315 hospitals.
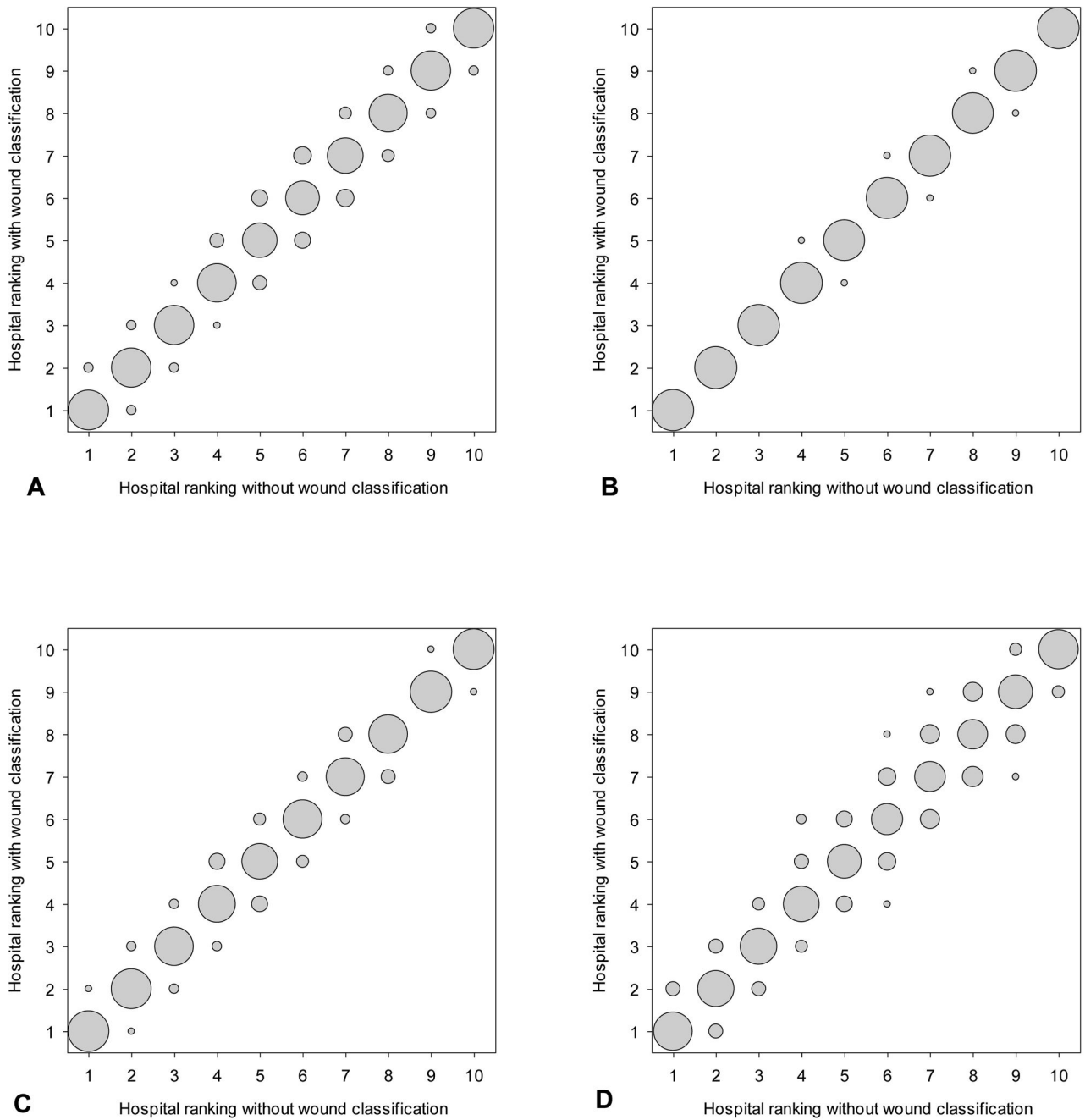
Figure 2.
Comparison in deciles for 315 hospitals in NSQIP (2011) using the model *excluding* wound classification and the model *including* wound classification for (A) any surgical site infection (SSI), (B) superficial incisional SSI, (C) deep incisional SSI, and (D) organ space SSI. Size of the bubbles represents proportion of hospitals: bigger bubble means more hospitals.

Table 1

30-day Postoperative Outcomes per Wound Classification in ACS NSQIP (2011)

30-d postoperative outcomes	Total		Wound classification							
			Clean		Clean/contaminated		Contaminated		Dirty/infected	
	n = 442,149	100.0%	n = 237,002	53.6%	n = 151,213	34.2%	n = 29,495	6.7%	n = 24,439	5.5%
Surgical site infection (SSI)	15,226	3.4%	4,279	1.8%	7,221	4.8%	1,658	5.6%	2,068	8.5%
Superficial incisional SSI	8,252	1.9%	2,785	1.2%	3,989	2.6%	824	2.8%	654	2.7%
Deep incisional SSI	2,483	0.6%	917	0.4%	948	0.6%	245	0.8%	373	1.5%
Organ space SSI	4,782	1.1%	593	0.3%	2,471	1.6%	639	2.2%	1,079	4.4%
Sepsis or septic shock	9,566	2.2%	2,439	1.0%	4,014	2.7%	1,205	4.1%	1,908	7.8%
Wound disruption	2,136	0.5%	708	0.3%	852	0.6%	224	0.8%	352	1.4%
Cardiac event *	3,227	0.7%	1,575	0.7%	913	0.6%	273	0.9%	466	1.9%
Venous thromboembolism	4,033	0.9%	1,651	0.7%	1,476	1.0%	341	1.2%	565	2.3%
Pneumonia	2,732	0.6%	1,040	0.4%	1,046	0.7%	227	0.8%	419	1.7%
Reintubation	4,552	1.0%	1,688	0.7%	1,691	1.1%	436	1.5%	737	3.0%
Prolong ventilation (>= 48 h)	5,602	1.3%	1,768	0.7%	1,731	1.1%	620	2.1%	1,483	6.1%
Urinary tract infection	7,520	1.7%	2,656	1.1%	3,665	2.4%	579	2.0%	620	2.5%
Renal failure	2,798	0.6%	1,004	0.4%	1,013	0.7%	259	0.9%	522	2.1%
Reoperation	15,528	3.5%	7,501	3.2%	4,752	3.1%	1,196	4.1%	2,079	8.5%
Mortality	5,764	1.3%	2,243	0.9%	1,601	1.1%	670	2.3%	1,250	5.1%
Length of stay, mean, median (IQR)	3.3	2.0 (0–4.0)	2.4	1.0 (0–3.0)	3.8	2.0 (1.0–5.0)	4.4	2.0 (1.0–5.0)	8.1	5.0 (2.0–10.0)

*
Cardiac event includes cardiac arrest requiring cardiopulmonary resuscitation and myocardial infarction.

IQR, interquartile range.

Table 2

Hospital Quality Comparisons in terms of Performance Correlation, Outlier Status Agreement, and Change in Rank between the Model Excluding Wound Classification Variable and the Model Including Wound Classification Variable in ACS NSQIP (2011)

| 30-d postoperative outcomes | Hospital quality comparisons | | | | |
| | Performance correlation | Outlier status agreement | Absolute change in RANK | | |
	R	Kappa	Mean	Median	Interquartile range
Surgical site infection (SSI)	0.9976	0.9508	4.5	3.0	(1.0, 6.0)
Superficial incisional SSI	0.9999	0.9872	0.9	1.0	(0.0, 2.0)
Deep incisional SSI	0.9978	0.9520	4.2	3.0	(1.0, 5.0)
Organ space SSI	0.9913	0.9007	9.9	6.0	(2.0, 14.0)
Sepsis or septic shock	0.9993	0.9845	2.3	2.0	(1.0, 3.0)
Wound disruption	0.9999	0.9907	0.9	1.0	(0.0, 1.0)
Cardiac event[*]	0.9999	0.9813	0.9	1.0	(0.0, 1.0)
Venous thromboembolism	0.9998	1.0000	1.6	1.0	(0.0, 2.0)
Pneumonia	0.9999	1.0000	0.7	0.0	(0.0, 1.0)
Reintubation	1.0000	1.0000	0.0	0.0	(0.0, 1.0)
Prolong ventilation (>= 48 hours)	0.9997	0.9877	0.7	1.0	(0.0, 3.0)
Urinary tract infection	0.9997	0.9832	1.7	1.0	(0.0, 3.0)
Renal failure	0.9998	1.0000	1.8	1.0	(0.0, 2.0)
Reoperation	0.9999	1.0000	1.1	1.0	(0.0, 2.0)
Mortality	0.9989	1.0000	3.0	2.0	(1.0, 5.0)

[*] Cardiac event includes cardiac arrest requiring cardiopulmonary resuscitation and myocardial infarction.

Table 3

Hospital Quality Comparisons in terms of Percentile and Decile between the Model Excluding Wound Classification Variable and the Model Including Wound Classification Variable in ACS NSQIP (2011)

30-d postoperative outcomes	Hospital quality comparisons							
	No. of hospitals that did not change in percentile (%of 315 total hospitals)	Absolute change in percentile			No. of hospitals that did not change in decile (% of 315 total hospitals)	Absolute change in decile		
		Mean	Median	Interquartile range		Mean	Median	Interquartile range
Surgical site infection (SSI)	40 (12.70%)	1.41%	0.95%	(0.32%, 1.90%)	261 (82.86%)	0.17	0.00	0.00
Superficial incisional SSI	138 (43.81%)	0.30%	0.32%	(0.00%, 0.63%)	309 (98.10%)	0.02	0.00	0.00
Deep incisional SSI	36 (11.43%)	1.33%	0.95%	(0.32%, 1.59%)	275 (87.30%)	0.13	0.00	0.00
Organ space SSI	25 (7.94%)	3.13%	1.90%	(0.63%, 4.44%)	219 (69.52%)	0.32	0.00	(0.0, 1.0)
Sepsis or septic shock	57 (18.10%)	0.74%	0.63%	(0.32%, 0.95%)	289 (91.75%)	0.08	0.00	0.00
Wound disruption	137 (43.49%)	0.28%	0.32%	(0.00%, 0.32%)	307 (97.46%)	0.03	0.00	0.00
Cardiac event[*]	136 (43.17%)	0.29%	0.32%	(0.00%, 0.32%)	303 (96.19%)	0.04	0.00	0.00
Venous thromboembolism	106 (33.65%)	0.51%	0.32%	(0.00%, 0.63%)	295 (93.65%)	0.06	0.00	0.00
Pneumonia	164 (52.06%)	0.23%	0.00%	(0.00%, 0.32%)	303 (96.19%)	0.04	0.00	0.00
Reintubation	175 (55.56%)	0.21%	0.00%	(0.00%, 0.32%)	305 (96.83%)	0.03	0.00	0.00
Prolong ventilation (>= 48 h)	92 (29.21%)	0.54%	0.32%	(0.00%, 0.95%)	293 (93.02%)	0.07	0.00	0.00
Urinary tract infection	84 (26.67%)	0.56%	0.32%	(0.00%, 0.95%)	299 (94.92%)	0.05	0.00	0.00
Renal failure	118 (37.46%)	0.40%	0.32%	(0.00%, 0.63%)	295 (93.65%)	0.06	0.00	0.00
Reoperation	131 (41.59%)	0.34%	0.32%	(0.00%, 0.63%)	305 (95.83%)	0.03	0.00	0.00
Mortality	60 (19.05%)	0.95%	0.63%	(0.32%, 1.59%)	285 (90.48%)	0.10	0.00	0.00

All values with p<0.0001.

[*] Cardiac event includes cardiac arrest requiring cardiopulmonary resuscitation and myocardial infarction.

SSI, surgical site infection.