TITLE PAGE

Project Title:

National Claims-Based Quality Measures for Surgical Site Infections

Principal Investigator:

Richard Platt, MD, MSc, Professor and Chair, Department of Population Medicine, Executive Director, Harvard Pilgrim Health Care Institute

Investigative Team:

Dale W. Bratzler, DO, MPH, Professor and Associate Dean, University of Oklahoma, Health Sciences Center

Michael S. Calderwood, MD, MPH, Assistant Professor of Medicine, Geisel School of Medicine, Dartmouth

Susan Huang, MD, MPH, Professor of Medicine, Medical Director, Epidemiology and Infection Prevention, University of California Irvine School of Medicine

Ken Kleinman, ScD, Associate Professor, Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst

Deborah Yokoe, MD, MPH, Professor of Clinical Medicine, University of California San Francisco

Organization:

Harvard Pilgrim Health Care, Inc.

Inclusive Dates of Project: 07/01/2012 – 04/30/2018

Acknowledgment of Agency Support: This project was made possible by Grant Number R18HS021424 from the Agency for Healthcare Research and Quality. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of AHRQ.

Grant Award Number: R18HS021424

STRUCTURED ABSTRACT

Purpose

This project expanded the use of claims-based methodology for SSI detection to determine its application for public reporting. This included the ability to address critical problems affecting current surveillance methods: non-standardized and incomplete detection and insufficient case mix adjustment.

Scope

There is a need for validated quality measures for inter-hospital comparisons of healthcare-associated infections, including SSIs. This applies to both major surgical centers and hospitals with lower procedure volumes. Claims data provide more comprehensive capture compared to traditional SSI surveillance by hospital infection control programs. Using Medicare data, we assessed claims-based approaches with the potential to be more uniform, complete, and inexpensive than conventional surveillance methods.

Methods
We assessed how well prior performance predicts future performance and evaluated the generalizability of Medicare claims-based SSI ranking to ranking using all-payer claims databases. Claims-based rankings were also compared to rankings based on hospitals' self-reporting to NHSN. We evaluated and resolved limitations on evaluating hospitals performing a low volume of procedures.
Results
Our claims-based method identified 89% of SSI following colon procedures and 81% of hysterectomy SSIs. Additionally, we demonstrated that operative hospital surveillance alone would have missed 7.2% of colon surgery and 13.4% of abdominal hysterectomy SSIs, further supporting the argument for use of large administrative datasets and applying validated diagnosis codes for SSI. Our work analyzing CMS hospital rankings for colon procedure SSIs showed that laparoscopic and open procedures were sufficiently different to require stratifying for fair inter-facility SSI comparisons.
Key Words
healthcare-associated infections, Medicare, performance measurement, hospital reimbursement, policy

PURPOSE

Approximately 300,000 surgical site infections (SSIs) occur each year in the U.S., with a cost of several billion dollars. Their prevention is a core measure of hospital performance for The Joint Commission. In addition, the Centers for Medicare and Medicaid Services (CMS) has implemented a value-based purchasing (VBP) program that targets SSI as a component of hospital quality. The VBP program will requires reporting of deep incisional and organ/space SSIs to CMS through the Centers for Disease Control and Prevention's (CDC) National Healthcare Safety Network (NHSN), and limits reimbursements to hospitals with poor performance.

There is particular interest in hospital reporting of SSI risk following coronary artery bypass graft (CABG) surgery, hip arthroplasty, and knee arthroplasty because of their high volume and cost. In addition, CABG and arthroplasty have been targeted by the Surgical Care Improvement Project (SCIP), a national partnership of public and private organizations formed in 2003 and dedicated to improving surgical care. Finally, the Agency for Healthcare Research and Quality (AHRQ) has developed SSI quality indicators that have been applied to administrative data collected through the Healthcare Cost and Utilization Project. The pursuit of hospital comparisons based upon SSI risk by CDC, CMS, Joint Commission, SCIP, AHRQ, and other national programs highlights the impact of these infections on healthcare burden, cost, and quality of surgical care.

Nevertheless, several problems hinder the inter-hospital comparison of SSI risk. First, the CDC definitions for SSI include subjective criteria, such as "diagnosis of [SSI] by the surgeon or attending physician." Second, hospitals vary greatly in the effort and resources they commit to SSI surveillance, leading to variable identification of SSIs, especially those occurring after discharge. Third, case mix adjustment across hospitals is limited to the data elements required by NHSN, which are resource intensive to collect. Fourth, there is not yet sufficient evidence that relative hospital performance is stable over time, such that current performance predicts future performance. And fifth, many hospitals perform a low volume of procedures, making their annual hospital-specific SSI risk too unstable for meaningful comparison.

Post-discharge surveillance of SSI accounts for the majority of SSIs and is most difficult to standardize. Both direct-to-patient outreach via phone calls and requests to surgeons to report SSIs have been shown to be ineffective.

Post-discharge surveillance is further complicated by the fact that patients may not return to the same hospital for subsequent infections. In fact, in a recent evaluation of the California administrative dataset using claims to detect SSI, we found that, on average, 20% of SSIs were identified in a hospital other than the original hospital performing the surgery. Equally important, among hospitals performing the initial surgery, the proportion of readmitted SSIs that returned to that index hospital ranged from 0-100%. Because Medicare and state-based claims allow linkage across hospitals, these can provide a more comprehensive method for SSI surveillance and be a critical resource for determining hospital performance and quality.

We sought to expand the use of claims-based methodology for SSI detection to determine its practical application for public reporting. We planned to do this in several ways:

First, we intended to expand the methodology to include colon surgery and abdominal hysterectomy, which have been targeted for SSI reduction both at the state and national levels. Mandated reporting on a national level by Medicare has caused these procedures to be relevant to all states and all hospitals, making these procedures the major focus of SSI public reporting and surgical care improvement.

Second, we refined all our claims-based detection methods to focus on serious (non-superficial) infections as specifically defined by CDC. In nearly all cases, legislative mandates have limited the reporting of SSI to CDC-defined deep incisional and organ/space SSIs due to their high cost and high morbidity. Similarly, although CDC's NHSN system requires reporting of all SSIs, CMS' VBP program focuses on only deep incisional and organ/space SSIs. We thus refined our CABG and arthroplasty claims-based surveillance methods to focus on deep incisional and organ/space SSIs. Using these methods, we further evaluated hospital rankings as a predictor of future performance. We used this to confirm or disprove the notion that identifying hospitals with good or poor past performance predicts better or worse care in the future. Such rankings could identify hospitals that might benefit from intervention.

Third, we evaluated whether the rank order of U.S. hospitals based on their Medicare claims SSI risk is similar to the rank order of hospitals when considering all patients. Medicare recipients account for approximately half of those undergoing CABG and arthroplasty procedures. We proposed to compare within-state hospital rankings in California and New York using Medicare claims to rankings generated by state-level administrative databases. Identification of similar outlier hospitals using Medicare and all-payer state databases would support the generalizability of hospital SSI rankings using Medicare claims alone. Otherwise, separate rankings may be required for a given subset of patients. In addition, we compared claims-based hospital rankings in California and New York to the SSI risks that hospitals report to NHSN in order to better understand how hospital rankings might differ using these two approaches for SSI surveillance.

Fourth, we explored the implications of public reporting and use of VBP on the large fraction of hospitals performing low numbers of surgical procedures each year. CDC excludes hospitals that perform fewer than 20 procedures per year from having procedure-specific SSI risks reported or compared to other hospitals. These hospitals may report their data to NHSN, but their SSI risk is not included when looking at national trends. This threshold of 20 procedures per year has yet to be supported, and there is a statistical basis to suggest that the choice of threshold and the ability to assess low volume hospitals could be improved. This is a critical area for exploration given evidence that many (20-30%) of U.S. hospitals perform fewer than 20 hip and knee arthroplasties annually.

In summary, our research was an innovative, comprehensive, and rigorous approach to SSI detection and ranking of U.S. hospitals. We showed how claims-based methods can address critical problems affecting current surveillance methods, including non-standardized and incomplete detection as well as insufficient case mix adjustment. It assessed how well prior performance predicts future performance. Moreover, this research evaluated the generalizability of Medicare claims-based SSI ranking to ranking using all-payer claims databases. Claims-based rankings were also compared to rankings based on hospitals' self-reporting to NHSN. Finally, it evaluated and resolved limitations on evaluating hospitals performing a low volume of procedures. These methods have the potential to be broadly used by states, as well as public and private payers, as a method for measuring the quality and outcomes of medical care.

SCOPE

The Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Control and Prevention (CDC), and the Centers for Medicare and Medicaid Services (CMS) have all expressed interest in developing validated quality measures for inter-hospital comparisons of healthcare associated infections, including SSIs. AHRQ's claims-based quality indicators have been applied to administrative data collected through the Healthcare Cost and Utilization Project. Medicare has begun implementation of a value-based purchasing program which will include deep incisional and organ/space SSIs reported by hospitals using the CDC's National Healthcare Safety Network (NHSN). Hospitals reporting higher SSI risks will receive lower reimbursement. Thus, there is a need for improved surveillance and validated quality measures targeting high-risk surgeries, both in major surgical centers and in hospitals with lower procedure volumes. Claims data have been shown to provide more standardized and comprehensive capture compared to traditional SSI surveillance by hospital infection control programs. We expanded our prior work on the use of claims data for SSI surveillance by targeting procedures mandated by states for public reporting and proposed for value-based purchasing by CMS.

This project responded to AHRQ's interest in patient safety and comparative effectiveness. Healthcare-associated infections (HAIs) are a major component of the AHRQ Patient Safety Portfolio and SSI is a key publicly reported HAI. As such, identifying appropriate and reproducible monitoring systems is critical for benchmarking and motivating best practice. Our work addressed the stage of AHRQ's patient safety research initiatives that focuses on the identification of SSI risks as a measure of important hazards associated with medical care. The interest in claims is fueled by the efficiency of using routinely collected national data and the advantage these data provide of comprehensively capturing healthcare utilization across all settings of care (hospitals, long term care, ambulatory settings, and home health).

In considering solutions to standardize SSI surveillance, our team has over 20 years of experience in assessing the value of administrative claims for identifying SSI throughout the spectrum of care delivery. Though claims-based detection does not work well for all diseases, we have repeatedly shown that use of claims data markedly increases identification of both pre- and post-discharge SSIs compared to traditional surveillance by hospital infection prevention programs.

Importantly, we also have shown that this type of surveillance can be implemented without the transfer of protected health information. Automated programs for this purpose have been developed and successfully used by national insurers on their administrative data. User-friendly instruction modules and data dictionaries to align data formats have enabled this surveillance method to be readily used to compare hospital performance based on claims-based indicators of SSI.

The success of this work has enabled a collaboration with CMS in pursuit of a claims-based method to rank hospitals by their SSI risk following the high volume and high cost procedures included in the SCIP program. We have worked closely with Dr. Dale Bratzler from the Oklahoma Foundation for Medical Quality, acting in its capacity as a national hospital quality resource center for Medicare's Quality Improvement Organization Program. Through this collaboration, we developed a nationally validated set of claims codes that perform well for detection of SSI indicators following CABG and arthroplasty procedures and improve the ability to rank U.S. hospitals by their SSI risk following these procedures. In this grant, we extend this process to colon and hysterectomy procedures. We have consistently found that this method is far more comprehensive for SSI detection compared to routine hospital surveillance.

Because of the recent focus of public reporting on deep incisional and organ/space SSIs, we sought to further demonstrate in this grant the ability of claims codes to better identify deep and organ/space SSIs compared to routine surveillance. Part of the goal was to identify a fair and comprehensive way that can trigger targeted chart review such that SSI detection is improved with a method that is labor-sparing for infection prevention programs.

Given the successful identification of SSIs using Medicare claims data for CABG and arthroplasty, there was further interest in this grant to confirm the relevance of Medicare claims data as a surrogate for claims data for all patients. Finally, we leveraged our validated claims-based code sets to analyze the effect of surgical procedure volume on SSI risk. The issue of whether low-volume hospitals can be appropriately assessed, and how, is important for all hospital-associated infection metrics, but especially surgical site infections, because many hospitals perform small numbers of procedures. The CDC's Healthcare Infection Control Practices Advisory Committee (HICPAC, a federal advisory committee) specifically mentions the instability of SSI risks reported from hospitals with low surgical volume. HICPAC does not, however, recommend a means of addressing this instability nor a clear definition of what constitutes low surgical volume. The CDC's National Healthcare Safety Network (NHSN) excludes hospitals that perform fewer than 20 procedures in a given year when reporting procedure-specific SSI risks; however, there is no basis provided for this threshold. CMS's Hospital Compare website does not report SSI risk for hospitals performing fewer than 25 procedures in a given year.

Aim #1: Validate a claims-based method to identify hospitals with unusually high risk of SSI following colon surgery and abdominal hysterectomy among Medicare beneficiaries. This builds on our prior use of Medicare claims to rank hospitals by SSI risk following CABG and hip arthroplasty.

The goal of this aim was to continue to provide a comprehensive, case mix adjusted method of identifying outlier institutions that might benefit from quality improvement interventions aimed at SSI reduction in high cost, high-volume procedures. Development of a claims-based quality measure will depend, in part, on the ability of such a measure to produce comprehensive standardized SSI detection regardless of variations in coding practices and its ability to perform reliably in the face of coding changes that might occur in response to value-based purchasing programs. We drew on our experience using Medicare claims to identify cases of SSI following CABG and hip arthroplasty to expand our work to target colon surgery and abdominal hysterectomy, common high-risk, high-volume surgical procedures. We tested a broad set of claims codes that are less sensitive to changes in coding practices over time or variability in coding practices between institutions.

This work expanded claims-based methodology for SSI detection other important high volume procedures. It provides a standardized method for benchmarking SSI risks that could be adopted by payers and health policymakers.

Aim #2: Develop and validate claims-based methods to predict hospitals' risk of serious SSIs following CABG, hip arthroplasty, colon surgery, and abdominal hysterectomy. This aim has two parts:
a. Determine whether these methods can identify deep incisional and organ space infections well enough to rank hospitals accurately.
b. Assess the ability of past performance, measured by claims, to predict future performance.

CMS, CDC, and AHRQ all target the subset of serious SSIs defined by the CDC's National Healthcare Safety Network (NHSN) system as deep incisional (at or beneath the fascial or muscle layer) and organ/space (any body area opened or manipulated during surgery excluding the skin incision, fascia, or muscle layers) in location. This subset of SSIs is associated with particularly significant preventable costs and morbidity. In general, deep incisional and organ/space SSIs are associated with a two- to four-fold relative risk of death during the post-operative hospitalization and a five- to six-fold relative risk of readmission within 30 days. Even amongst those who are successfully treated for these SSIs, there is a 60% increase in the need for ICU care. Among patients 65 and older, SSIs are associated with a three-fold increase in length of stay and a two-fold increase in hospital charges. Thus, reductions in hospital risks of deep incisional and organ/space infections should significantly lower hospital expenditures while improving morbidity and mortality in the postoperative period.
In this aim, we developed and validated a specific set of claims codes that targets the identification of deep incisional and organ/space SSI following CABG, hip arthroplasty, colon surgery, and abdominal hysterectomy. These procedures are the most commonly targeted procedures by state and national mandates, and they represent three of the highest-volume, highest-cost procedures in the U.S.

Aim #3: Compare hospital SSI rankings using Medicare claims to rankings using a) comprehensive all-payer claims from California and New York and b) SSIs identified and reported by California and New York hospitals to CDC's NHSN system.

While Medicare pays for over 50% of CABG and arthroplasty procedures in U.S. hospitals, all payers and state departments of public health have an interest in identifying hospitals that are outliers for SSI performance. The previous aims have discussed the development and validation of claims-based surveillance tools for identifying best performing and worst performing hospitals using Medicare claims. In this aim, we evaluated the generalizability of Medicare rankings versus those generated by the same code lists in all payer databases from California and New York. We also compared claims-based SSI rankings to self-reported data that these hospitals provide to the CDC/NHSN national surveillance systems. Understanding similarities and differences between these systems had the potential to elucidate the value of using Medicare and all-payer claims as an efficient way to identify outlier hospitals.

Aim #4: Assess the impact of surgical procedure volume on hospital-specific SSI risk. This aim has two components:
a. Assess whether low-volume hospitals perform differently (better or worse) than high-volume hospitals.
b. Develop methods to monitor the performance of hospitals that perform a low volume of procedures.

Value-based purchasing is based on the assumption that high performing hospitals will continue to provide high-quality care and that financial reimbursement for high-quality care will encourage improvement in poorly performing hospitals. This assumption is predicated on annual reporting metrics that are sufficiently meaningful to judge the quality of care. However, the annual SSI risk measure in small volume hospitals is almost certainly too imprecise because of random variation to be useful for this purpose. For example, 18% of hospitals perform 10 or fewer knee arthroplasties annually on Medicare patients. A single SSI in a hospital that performs 10 procedures would shift its annual SSI risk from 0% to 10%. The 0% risk would classify the hospital as a best performing hospital, and a 10% risk would deem it a worst performing hospital.

A structured and comprehensive assessment of meaningful performance measures for low volume hospitals is critically needed for several reasons. First, there are thousands of low-procedure-volume hospitals, constituting a substantial proportion of the total. Second, adverse events may be higher in hospitals performing fewer procedures (described below). Thus, the ability to reliably identify poor performance may be especially important in this group. These hospitals may be important venues for targeted intervention to improve outcomes. Third, the existence of different, arbitrary thresholds in current national reporting is confusing and problematic. Development of statistically grounded measures is important for meaningful inter-hospital comparisons. Thresholds are likely not to be one-size-fits all but rather depend on the estimated procedure-specific risk of SSI. Finally, the solution may not be threshold-based. Alternative statistical methods may reveal better methods to assess low-volume hospitals.

METHODS

Aim #1: We conducted two separate retrospective cohort studies of Medicare beneficiaries, one group that underwent colon surgery and one group that underwent abdominal hysterectomy in U.S. hospitals between 10/1/12 and 12/31/13. These study cohorts will include fee-for-service Medicare beneficiaries who are ≥65 years old at the time of surgery.

In preliminary work in two academic hospitals associated with the CDC Prevention Epicenters program, we identified six ICD-9 procedure codes plus 14 ICD-9 diagnosis codes for colon surgery and six ICD-9 diagnosis codes for abdominal hysterectomy that captured >90% of the SSIs identified by traditional surveillance with significant increases in overall case detection above those previously identified.

For each patient included in our cohorts, we screened all diagnosis and procedure codes submitted under Medicare Part A from inpatient facilities. We identified patients with any of our codes suggestive of SSI during the admission when the surgery occurred or on readmission to the hospital where the surgery occurred within 30 days. These are cases that a hospital should be reasonably expected to identify.

Our ranking approach also utilized CMS administrative data on age at the time of surgery, gender, and Romano score for each patient. The Romano score is a comorbidity index based on ICD-9 diagnosis codes present in claims in the prior year. It assigns weights to different health conditions and has been revised to predict mortality in Medicare patients. Higher scores predict more comorbid illnesses. Our prior work showed this score is highly predictive of SSI following CABG.

Our SAS programs returned an output of descriptive characteristics of the patient cohort as well as hospital-specific descriptors such as the number of colon surgery and abdominal hysterectomy procedures performed and the median age and comorbidity score of surgical candidates.

Using claims codes as indicators of potential SSI, we calculated an unadjusted claims-based SSI risk for each U.S. hospital based on the proportion of colon surgery and abdominal hysterectomy procedures found to have ≥1 code of interest within 365 days. We then ranked hospitals by their adjusted SSI risk obtained from a generalized linear mixed model using patient-level data as in our previous work. This model adjusted for age, gender, and Romano score while accounting for clustering of SSI at the hospital level. The modeled outcome was the presence of any of the diagnosis or procedure codes of interest. The mixed effects model could then be used to derive hospital-specific adjusted SSI risk.

To validate the ranking of hospitals by claims-based diagnosis and procedure codes, we reviewed full-text medical records on a random sample of 150 patients with one of these codes from best decile hospitals and another sample of 150 patients with one of these codes from worst decile hospitals. Records were evaluated for evidence of chart-confirmed SSI using CDC/NHSN criteria.

Once chart reviews were complete, we calculated the fraction of charts with a diagnosis or procedure code suggestive of SSI that were confirmed by review using CDC/NHSN criteria. We then performed logistic regression for the outcome of chart-confirmed SSI using independent variables of age, gender, comorbidity score, and whether the surgery was performed in a best versus worst decile hospital based upon claims rankings. Cases were defined as patients with at least one of the 65 diagnosis or procedure indicator codes for SSI in which medical record confirmed SSI, and non-cases will be defined as patients who were determined on review of the full-text record not to meet the CDC/NHSN criteria for SSI. Based on the high sensitivity of our selected claims codes for detecting SSIs, we assume that patients without one of these codes do not have an SSI. Notably, our described pilot study had confirmed 100% sensitivity of identifying SSI, twice the sensitivity of routine hospital surveillance.

The logistic regression allowed us to calculate an adjusted odds ratio comparing the SSI risk in hospitals ranked in the worst performing 10% of hospitals compared to hospitals in the best performing 10%. An odds ratio greater than one confirms that these hospitals, identified through claims, do in fact have a higher probability of chart-confirmed SSI.

Aim #2: For the development and validation of claims-based code sets to detect deep incisional and organ/space SSIs, we conducted four retrospective cohort studies of Medicare patients undergoing CABG, hip arthroplasty, colon surgery, and abdominal hysterectomy. These were secondary analyses on prior national validation datasets for these procedures. The CABG procedures were from 2005, the hip arthroplasties were from 2005-2007, and both the colon surgeries and abdominal hysterectomies were from 2012-2013.
To evaluate the ability of past performance to predict future performance, we assessed the stability of hospital SSI risk and outlier status using State Inpatient Databases containing claims data in inpatient discharge abstracts. For this part of the aim, we examined procedures from 2005-2010, a 6-year span.

For each of the four procedures, we identified the subset of patients with deep incisional or organ/space SSIs confirmed on chart review and found ICD-9 codes used at least once in their claims. Each procedure was assessed independently. Ordering the codes from highest to lowest sensitivity, we assessed the cumulative capture of each additional code.

Procedure-specific claims-based code sets were applied to national data from Medicare beneficiaries undergoing CABG, hip arthroplasty, colon surgery and abdominal hysterectomy in 2010. We provided a SAS program to our CMS collaborators to select patients for chart review. Patients who had another Surgical Care Improvement Project (SCIP) procedure on the day of the index surgery or in the 60 days before surgery will be excluded in order to reduce uncertainty in attributing an SSI to a specific procedure.

We also excluded patients with a diagnosis or procedure code suggestive of infection at the surgical site on the day of surgery or in the 30 days prior to surgery. SSI outcomes based upon these SSI indicators were assessed for the 365 days following each procedure, in accordance with CDC's National Healthcare Safety Network (NHSN) surveillance criteria for surgeries with an implant.

For each procedure, we ranked hospitals based upon the proportion of patients who had a claims code suggestive of SSI within 365 days of surgery. We performed both unadjusted rankings, as well as adjusted rankings derived from a generalized linear mixed model using patient-level data. The modeled outcome was the presence of any of the diagnosis or procedure codes of interest. Similar to our prior work, this model adjusted for age, gender, and Romano score while accounting for clustering of SSI at the hospital level. The mixed effects model was then used to derive hospital-specific adjusted SSI risk.

Outlier hospitals were defined as those in the top or bottom decile of the adjusted rankings for deep incisional and organ/space SSI. In a process identical to Aim 1, a random sample of patients was selected for chart review from patients undergoing each procedure who have a claims-based SSI indicator within 365 days of surgery. Charts were reviewed using CDC NHSN criteria for deep incisional or organ/space SSI only.

Following chart review, we determined whether the sets of codes are equally effective in identifying deep incisional and organ/space SSIs in the best and worst performing 10% of hospitals. To do this, we assessed whether each individual code and the set of codes differentially predicts SSI in the respective deciles. Significance was assessed using chi-squared tests at an alpha = 0.05.

Logistic regression was used for the outcome of chart-confirmed deep incisional or organ/space SSI using the same case and non-case definitions that were used in Aim 1. Independent variables included age, gender, Romano comorbidity score, and whether the surgery was performed in a best versus worst decile hospital based upon adjusted claims rankings. This allowed us to calculate an adjusted odds ratio comparing the risk of deep incisional or organ/space SSI in hospitals ranked in the worst-performing 10% compared to hospitals in the best-performing 10%. An odds ratio greater than one would confirm that the best claims-identified hospitals do in fact have a lower probability of chart confirmed SSI than the worst claims-identified hospitals.

For procedures with nationally validated claims-based code sets for detection of deep incisional and organ/space SSI, we assessed the stability of adjusted hospital SSI risks and rankings for each year between 2005 and 2012. For each of the validated procedures, we fit hospital-level dynamic generalized mixed models. For modeling risks, each hospital's risk in a given year was assessed as a predictor of its risk in the following year in a dynamic linear mixed model. For the rankings, we fit a dynamic generalized logistic mixed model with each hospital's decile status (three categories: best decile, worst decile, and all other middle deciles) in a given year as a predictor of the next year's status similarly trichotomized. In these models, the SSI risk (or decile status) in the first year and the last year of the study period were used only once, as a predictor and as an outcome, respectively. Each intermediate year's risk (or decile status) was used twice: once as an outcome and again as a predictor. The model was fit as a repeated measures model, with N-1 repeated measures made up out of the N years of observations. The net result for the SSI risks was the proportion of the following year's risk that is the result of the prior year, as well as the proportional reduction in variability when using the prior year as a predictor, relative to not using it. The latter is similar to the R^2 measure from a simple linear regression.

For the trichotomized ranks, the model result was the odds ratio that a hospital receiving a worst decile designation in 1 year will receive it in the next as well, versus being in the middle deciles. The model also generated the odds that a hospital in the best decile will retain that ranking the next year, relative to being in the middle deciles. The odds of moving from one extreme to the other was also calculated. Standard methods for generating standard errors and confidence intervals or p values were applied. We also performed sensitivity analyses assessing whether longer periods of prior performance better predicts future performance.

Aim #3: Compare hospital SSI rankings using Medicare claims to rankings using a) comprehensive all-payer claims from California and New York and b) SSIs identified and reported by California and New York hospitals to CDC's NHSN system.

We used these two large mandatory state hospital discharge databases that include all-payer administrative claims data to compare state-specific SSI rankings using Medicare data alone to SSI rankings based upon all-payer, all-patient databases. We specifically evaluated the comprehensive inpatient databases from California and New York, two states that retained an encrypted identifier that allows claims-based tracking across multiple facilities. The patient population included all adults in these two states who underwent CABG, hip arthroplasty, or knee arthroplasty from January 2010-December 2012 as well as the entirety of 2014. The latter 2014 evaluation enabled us to assess the performance of our claims codes following the conversion from ICD-9 to ICD-10. Reliable ICD-10 coding was required by October of 2013.

Our goal was to assess whether these two approaches identify the same outlier hospitals. We focused on the comparison of deep incisional and organ/space SSIs due to the interest at the national (CMS value-based purchasing) and state level for targeting these high cost and high morbidity events. We also focused on deep incisional and organ/space SSIs because the mandatory state databases did not include outpatient visits, and this subset of SSIs is highly likely to require inpatient care.

Using these states' all-payer databases, which included patients lacking health insurance, we have assessed the rankings of all hospitals in the state using our ICD-9 and CPT SSI claims codes developed and validated to detect patients with deep incisional and organ/space SSI. We compared rankings based on all-payer claims data to those based upon Medicare claims alone. We also compared rankings based on Medicare claims versus non-Medicare claims found in these states' all-payer databases. All SSIs were evaluated through 365 days post-procedure, consistent with the CDC/NHSN criteria for surgeries with an implant. In the case of CABG, sternal wires were considered an implant, and in the case of arthroplasty, the joint replacement was an implant. Similar to Aim 1, we excluded patients who had another major procedure on the day of the index surgery or within 60 days before surgery. We defined major procedures as those targeted in the national Surgical Care Improvement Project (SCIP), which include CABG, other cardiac surgery, hip arthroplasty, knee arthroplasty, vascular surgery, colon surgery, and hysterectomy. Patients with multiple surgical dates for the same SCIP procedure during their index hospitalization were also excluded. Finally, we excluded patients with a diagnosis or procedure code suggestive of infection at the surgical site on the day of surgery or within 30 days prior to surgery.

Similar to Aim 2, we ranked California and New York hospitals based on SSI risk determined by the proportion of procedures with a deep incisional and organ/space SSI indicator code adjusted for age, gender, and Romano comorbidity score. Rankings were derived from a random effects matrix generated from a regression model predicting SSI claims indicators while adjusting for clustering by hospital.

We have produced a risk-adjusted ranking of all California and New York hospitals for each of three procedures by their risk of deep incisional and organ/space SSI.

Through our collaboration with the Oklahoma Foundation for Medical Quality, acting in its capacity as a national hospital quality resource center for Medicare's Quality Improvement Organization Program, we separately analyzed Medicare claims restricted to California and New York hospitals. We then conducted retrospective cohort studies on all Medicare beneficiaries who underwent CABG, hip arthroplasty, or knee arthroplasty during the study periods, screening for claims suggestive of deep incisional and organ/space SSI and generating risk-adjusted rankings of all California and New York hospitals, as previously described. To align with the data that is available in state inpatient databases, we included Medicare Part A inpatient claims only. Nevertheless, we used our validated 2010 chart reviewed data from Aim 2 to calculate and report the sensitivity of Part A claims to capture all deep incisional and organ/space SSI identified by applying our code sets to both inpatient and outpatients claims (Medicare Part A and B).

For each year, we had three adjusted rank orders for California and New York hospitals for each procedure. The first ranking used Medicare claims only, the second ranking used non-Medicare claims only, and the third ranking used all-payer data. Our goal was to assess whether hospital rankings using non-Medicare and all-payer data are similar to hospital rankings using Medicare data alone, and whether this similarity is preserved year to year. If these data sources similarly identified outlier hospitals, it would allow Medicare data to be broadly used as a trigger for further assessment of hospitals that may need improvement.

In addition, we compared our claims-based SSI rankings of California and New York hospitals to hospital-specific rankings based upon self report for CABG, hip arthroplasty, and knee arthroplasty procedures. These self-reported risks are provided by hospitals to the CDC/NHSN system under state legislative mandate and then made publicly available through the state departments of public health. Because California's public reporting mandate for SSI encompasses surgical procedures performed from mid-2011 onward, we used the entirety of 2012 and 2014 for this endeavor. Similar to the above, the inclusion of 2014 enabled us to assess the performance of our claims codes following their conversion from ICD-9 to ICD-10. Using California's publicly reported NHSN SSI data, we compared rankings based on claims-based indicators of deep incisional and organ space SSIs for all three procedures: CABG, hip arthroplasty, and knee arthroplasty. Using New York's publicly reported NHSN SSI data, we compared rankings based on claims-based indicators of deep incisional and organ space SSIs for the two mandated procedures: CABG and hip arthroplasty.

We ranked each hospital based on the estimated SSI risk, separately within each data set. Then, we measured the absolute value of the difference in ranks. This comparison was done for Medicare versus non-Medicare claims as well as for Medicare versus all-payer claims. If the median difference in rank value is less than 3% of the sample size, we considered this evidence of good agreement. So, for example, if there are 100 hospitals, we would require that at least half the hospitals are ranked within three places of their ranking within comparator data sets.

Aim #4: Assess the impact of surgical procedure volume on hospital-specific SSI risk. This aim has two components:
a. Assess whether low-volume hospitals perform differently (better or worse) than high-volume hospitals.

b. Develop methods to monitor the performance of hospitals that perform a low volume of procedures. Current quality reporting systems ignore 20-30% of hospitals because of low procedure volumes.

We used 2005–2011 Medicare Provider Analysis and Review Research Identifiable Files to identify all short-stay acute care US hospitals performing CABG and primary hip arthroplasty on fee-for-service Medicare patients between January 1, 2005, and December 31, 2011, based on Medicare Part A inpatient claims data.

CABG cases and primary hip arthroplasty cases were identified using International Classification of Diseases, Ninth Revision (ICD-9) codes. We then assessed claims within 90 days of the surgical procedure for ICD-9 codes suggestive of a deep and organ/space SSI following CABG and following primary hip arthroplasty. To focus on surgical complications rather than pre-existing infections, we excluded all codes listed as present on admission at the time of surgery while including present on admission codes during any readmissions to a hospital within 90 days of the surgical procedure. For patients who underwent another major surgery in the 90-day postoperative surveillance window, we censored our surveillance at the time of the subsequent surgery. We used 2012 Medicare Provider Analysis and Review Research Identifiable Files data to capture coding and readmissions for procedures performed in the last 90 days of 2011.

Medicare procedure volume was categorized into volume categories of surgical cases on Medicare patients each year, and hospitals could change volume categories across years. Categories corresponded to hospitals that would be excluded from public reporting based on expected SSI rates of 4%, 2%, 1%, 0.05%, respectively, using the current CMS methodology of excluding hospitals with <1 expected SSI in a given year based on procedure volume. We used logistic regression to calculate the odds of having an SSI code by the annual Medicare surgical volume. Generalized estimating equations were used to control for repeated measures across years within individual hospitals, and the model was adjusted for the age, sex, and comorbidities of each Medicare patient. We also examined the relationship between continuous procedure volume and the probability of a patient having an SSI code graphically using generalized additive models.

Within each year, we fit a logistic regression mixed effects model with random intercepts for each hospital, including each individual's age, sex, and coded comorbidities as fixed effects. The predicted random intercepts from the models for each year indicate case mix-adjusted relative performance. Hospitals were ranked based on these predicted random intercepts and divided into case mix-adjusted quartiles of performance. We then used logistic regression, with generalized estimating equations to account for repeated measures across years within hospital, to model the outcome of being in the worst quartile next year based on (1) worst, middle two, and best quartile status this year, for all hospitals combined and by volume category, and (2) worst, middle two, and best quartile status this year and the year prior by volume category. Our modeling of quartile status was based on the fact that the CMS HAC Reduction Program uses quartiles to determine financial penalties.

Additional analyses were performed evaluating hospital SSI rates and rankings following colon surgery in fee-for-service Medicare beneficiaries between January 1, 2009, and November 30, 2013, based on Medicare Part A inpatient claims data. Using individual-level logistic regression models, we assessed the odds of SSI in categories of procedure volume while accounting for case mix adjustors using age, gender, and comorbidity scores, which have been found to be highly significant in prior work.

These models will use mixed effects methods to account for clustering within hospital and will further allow different random effect variance depending on category.

Adjusted SSI rates were calculated for each hospital's overall colon procedures as well as those stratified by laparoscopic and open status. Concomitant intra-abdominal procedures for each instance of colon surgery were identified by date-matched ICD-9 codes. Adjusted SSI rates were derived from generalized linear mixed models accounting for age, sex, race, Elixhauser comorbidity score, concomitant colon and non-colon intra-abdominal procedures, and laparoscopy (for overall colon procedures only) while accounting for clustering at the hospital level. Adjusted odds ratios for each variable were converted to an overall hospital adjusted SSI rate based upon the characteristics of each hospital's patients undergoing the procedures in reference to the mean SSI rate across hospitals. We ranked hospitals by their adjusted SSI rate and also grouped them into deciles by SSI rates for total colon procedures. This was also done separately for laparoscopic and open procedure strata. We compared the rank order and decile assignment between overall and stratified groups.

RESULTS

Aim #1: Validate a claims-based method to identify hospitals with unusually high risk of SSI following colon surgery and abdominal hysterectomy among Medicare beneficiaries. This builds on our prior use of Medicare claims to rank hospitals by SSI risk following CABG and hip arthroplasty.

We found that six ICD-9 codes for identifying SSI following abdominal hysterectomy detected 1.7 times as many infections as routine surveillance. The overall SSI rate utilizing claims-based screening followed by medical record review (claims-enhanced surveillance) was 2.9% versus 1.7% based on routine surveillance. Using claims codes to trigger chart review identified 1 SSI for every 2 hysterectomy procedures reviewed. The negative predictive value (NPV) of the selected codes was 98.5%.

The 20 ICD-9 codes for identifying SSI following colon surgery detected 3.9 times as many infections as routine surveillance. The overall SSI rate utilizing claims-enhanced surveillance was 8.7% versus 2.2% based on routine surveillance. Using claims codes to trigger chart review identified 1 SSI for every three colorectal procedures reviewed. The NPV of the selected codes was 97.7%.

We requested records on 600 colorectal patients in 289 U.S. hospitals and 515 hysterectomy patients in 200 U.S. hospitals. We received sufficient records to determine an outcome in 550/600 colorectal cases (92%) and 466/515 hysterectomy cases (90%). We identified 140 colorectal SSIs (45 superficial SSIs, 95 deep and organ/space SSIs) and 127 hysterectomy SSIs (58 superficial SSIs, 69 deep and organ/space SSIs). SSI diagnosis codes had a sensitivity of 89% for colorectal cases and 81% for hysterectomy cases. The PPV of SSI codes was 38% for colorectal cases and 61% for hysterectomy cases. This corresponds to one SSI identified for every 2.6 records reviewed for colorectal cases and 1.6 records reviewed for hysterectomy cases. Reviewing records without an SSI diagnosis code captured some additional cases, but the number of SSIs identified per records reviewed significantly declined. Review of cases missed by current IQR codes did not reveal additional infection codes that would improve sensitivity.

Aim #2: Develop and validate claims-based methods to predict hospitals' risk of serious SSIs following CABG, hip arthroplasty, colon surgery, and abdominal hysterectomy.

Our analyses included 143 patients with deep and organ/space surgical site infections (D/OS SSI) after CABG and 175 patients with D/OS SSI after hip arthroplasty. For CABG, nine International Classification of Diseases, Ninth Revision (ICD-9) diagnosis codes identified 92% of D/OS SSI, with one D/OS SSI identified for every four cases with a diagnosis code. For hip arthroplasty, six ICD-9 diagnosis codes identified 99% of D/OS SSI, with 1 D/OS SSI identified for every two cases with a diagnosis code.

Additional analyses have shown among 60,059 colon surgeries at 285 hospitals and 64,918 abdominal hysterectomies at 270 hospitals, 5,921 (9.9%) colon surgeries and 1,481 (2.3%) abdominal hysterectomies received a diagnosis code for SSI within the 30 days following surgery. Operative hospital surveillance alone would have missed 7.2% of colon surgery and 13.4% of abdominal hysterectomy SSIs. The proportion of an individual hospital's SSIs detected during hospitalizations at other hospitals varied widely. Including nonoperative hospital SSIs resulted in improved relative ranking of 11 (3.9%) colon surgery and 13 (4.8%) hysterectomy hospitals so that they were no longer in the worst performing quartile, mainly among hospitals with relatively high surgical volumes.

In an addition to finding claims-based methods for assessing SSI risk of colon surgery and abdominal hysterectomy, we collaborated with the California Department of Public Health (CDPH) to validate these methods in a large set of California hospitals. Upon validation review, Infection Preventionists at the CDPH identified 239 SSIs following colon surgery at 42 hospitals and 76 SSIs following abdominal hysterectomy at 34 hospitals. For colon surgery, traditional surveillance had a sensitivity of 50% (47% for deep incisional or organ/space [DI/OS] SSI), compared to 84% (88% for DI/OS SSI) for claims-based surveillance. For abdominal hysterectomy, traditional surveillance had a sensitivity of 68% (67% for DI/OS SSI) compared to 74% (78% for DI/OS SSI) for claims-based surveillance. Claims-based surveillance was also efficient, with one SSI identified for every two patients flagged for review who had undergone abdominal hysterectomy and for every 2.6 patients flagged for review who had undergone colon surgery. Overall, CDPH identified previously unreported SSIs in 74% of validation hospitals performing colon surgery and 35% of validation hospitals performing abdominal hysterectomy.

This work had demonstrated that while CDC SSI surveillance definitions are standardized, SSI case findings differ across hospitals and fail to identify one third to one half of all SSIs. Because readmission, reoperation, and positive microbiology are key triggers for traditional surveillance, cases in which these factors are absent provide good examples of when claims-based surveillance can augment traditional methods. Requiring that postoperative billing codes (ICD-9/ICD-10 diagnosis codes) be incorporated into routine surveillance will improve the sensitivity and efficiency of case finding and will improve the validity of publicly reported comparisons of hospital surgical performance. We demonstrated that this methodology can be effectively used for external validation of publicly reported data.

Moreover, our validated claims-based methods to predict hospitals' risk of serious SSIs was used as a secondary outcome to a national trial of a SSI prevention bundle for hip and knee arthroplasty. In the assessment of a multi-state quality improvement campaign (Project JOINTS) that used the Institute for Healthcare Improvement's (IHI) Rapid Spread Network (RSN) to promote adoption of evidence-based surgical site infection (SSI) prevention practices. In the analysis of the change in SSI outcomes in five different Project JOINTS intervention states, postoperative Medicare claims data for International Classification of Diseases, Ninth Revision (ICD-9) codes suggestive of an SSI within 90 days of the surgical procedure were used to identify SSI events.

Our claims-based methods had been used to demonstrate that this quality improvement campaign was effective in reducing SSIs following hip and knee arthroplasty and that, compared to a set of matched comparison states, there was a significant reduction in SSIs in intervention states.

Aim #3: Compare hospital SSI rankings using Medicare claims to rankings using a) comprehensive all-payer claims from California and New York and b) SSIs identified and reported by California and New York hospitals to CDC's NHSN system.

We first evaluated hospital performance based on Medicare versus Commercial Payer claims data for CABG SSI outcomes (IDWeek 2014, Abstract 915), analyzing data from 119 California and 41 New York hospitals that performed CABG on Medicare-insured and commercially insured patients between 2009 and 2011. We used Pearson correlations to compare risk-adjusted odds of SSI for each hospital based on outcomes in Medicare-insured patients versus commercially insured patients. These correlations were better in New York than in California, with some improvement in California when limited to hospitals performing 50+ procedures in both groups. This may be due to a large proportion of relatively low-volume hospitals (<100 annual CABG procedures) in California compared with New York due to the fact that New York has a Certificate of Need law requiring that hospitals licensed to perform cardiac surgery maintain an annual minimum of 100 procedures on adult patients. Overall, though, there is a moderate-to-strong correlation in the adjusted odds of surgical site infection in Medicare-insured versus commercially insured patients undergoing CABG at the same hospital.

Over a 3-year period (2009-2011), we compared outcomes in adult (18+) patients with Medicare as their primary insurance vs. Commercial Payer as their primary insurance, looking at the correlations comparing a hospital's SSI outcomes for Medicare patients vs. the same hospital's SSI outcomes for Commercial Payer patients. The Pearson correlation comparing outcomes between these two groups in New York hospitals was strong (0.70) but only moderate for California hospitals (0.51). Of note, however, if we limited analysis to hospitals performing an average of 50+ procedures per year on Medicare patients and 50+ procedures per year on Commercial Payer patients, the correlation for California hospitals increased to 0.68 and remained strong (0.76) for New York hospitals. This may be due to a large proportion of lower volume hospitals in California compared with New York, because New York has a Certificate of Need law requiring that hospitals licensed to perform cardiac surgery maintain an annual minimum of 100 procedures on adult patients.

After extensive evaluation, we have found sufficiently inconsistent results that we are unable to explain and have ceased pursuing this exploration.

Aim #4: Assess the impact of surgical procedure volume on hospital-specific SSI risk.

Our analyses showed that case-mix adjusted SSI risk based on claims was highest in hospitals performing <50 CABG/year and <200 hip arthroplasty/year compared with hospitals performing ≥200 procedures/year. Hospitals performing fewer than 50 procedures per year accounted for more than a quarter of all U.S. hospitals performing CABG procedures from 2005 to 2011 while only accounting for 7%–8% of the surgical volume. Patients undergoing these procedures at hospitals performing <50 Medicare procedures per year had significantly higher odds of an SSI code: 30% higher for patients in hospitals performing fewer than 25 procedures and 21% higher in hospitals performing 25-49 procedures.

For hip arthroplasty, there was an inverse correlation between procedure volume and patients' SSI risk, with patients at hospitals in each of the four lower-volume categories having significantly higher risk than patients at hospitals performing ≥200 procedures per year. The excess odds of an SSI code ranged from 58% in hospitals performing fewer than 25 procedures to 14% in hospitals performing 100-199 procedures. Over 90% of US hospitals performing hip arthroplasty procedures from 2005 to 2011 had SSI risks that were significantly higher than those in the largest-volume hospitals.

At that same time, hospitals in the worst quartile in a given year based on claims had a low probability of remaining in that quartile the following year. This probability increased with volume, and when using 2 years' experience, but the highest probabilities were only 0.59 for CABG (95% confidence interval, 0.52–0.66) and 0.48 for hip arthroplasty (95% confidence interval, 0.42–0.55).

We determined that aggregate SSI risk is highest in hospitals with low annual procedure volumes, yet these hospitals are currently excluded from quality reporting. Our data suggest that procedure volume is a strong predictor of outcome following CABG and hip arthroplasty, with the highest SSI risk found in those hospitals currently excluded from public scrutiny. In addition, though SSI rates have declined nationally over time, our modeling of longitudinal data from individual hospitals suggests that annual SSI performance rankings may be highly unstable from year-to-year, even for hospitals performing many procedures. This raises important concerns about the use of SSI data for both quality reporting and reimbursement programs.

As for why coded SSI rates declined with increasing surgical volume only up to around 100 annual CABG cases on Medicare fee-for-service patients, though the volume outcome relationship continued beyond this surgical case volume for hip arthroplasty, we do not know of existing literature highlighting this finding. It is not clear why this should differ between CABG and hip arthroplasty, and it is worth further investigation.

In exploring the impact of a broader sampling strategy, we did find a modest improvement in prediction by using quartile ranking over 2 consecutive years to predict whether a hospital would remain in the worst quartile. Even with this extra year of data, however, prediction remained poor. Even for higher-volume hospitals, year-to-year random variation makes past experience an unreliable estimator of current performance.

Our additional analyses related to this aim have found that other characteristics of colon surgery which affect stability, such as those undergoing laparoscopic vs. open procedures, should not be grouped together for quality reporting and reimbursement programs. We demonstrated a need to stratify these unique populations for fair inter-facility SSI comparisons. In an analysis of 694,818 colon procedures performed by 4,094 hospitals between January 1, 2009, and November 30, 2013, laparoscopic and open adjusted SSI rates varied significantly across hospital deciles. When stratified by laparoscopic and open procedures, correlation by individual hospital and decile rankings, respectively, were both poor ($r=0.24$ and $r=0.26$, $p<0.001$).

A manuscript detailing this work has been drafted and will be submitted to Clinical Infectious Diseases for publication.

LIST OF PUBLICATIONS and PRODUCTS

Abstracts and Presentations

Calderwood MS, Kleinman K, Canning C, et al. Improving Surveillance for Deep and Organ/Space Surgical Site Infections Following Coronary Artery Bypass Grafting and Hip Arthroplasty. IDWeek. San Francisco, CA. October 2013.

Calderwood MS, Huang SS, Keller V, et al. Validation of Surgical Site Infection Data in California Hospitals Demonstrates Variable Case Identification Following Abdominal Hysterectomy. IDWeek. Philadelphia, PA. October 2014.

Calderwood MS, Huang SS, Keller V, et al. Use of Medicare Claims Data to Identify Cases of Surgical Site Infection Following Colon Surgery Identified Many Unreported Infections in a State-Wide Validation. IDWeek. Philadelphia, PA. October 2014.

Calderwood MS, Kleinman K, Murphy MV, et al. Is a Hospital's Surgical Site Infection Rate Among Medicare-Insured Patients a Good Indicator of Outcome for Commercially-Insured Patients? IDWeek. Philadelphia, PA. October 2014.

Calderwood MS, Kleinman K, Murphy MV, et al. U.S. Hospitals with Low Surgical Volume are Excluded from Public Reporting but have Higher Rates of Surgical Site Infection. IDWeek, Philadelphia, PA. October 2014.

Calderwood MS, Yokoe DS, Murphy MV, et al. State-Level Quality Improvement Campaigns Can Successfully Reduce SSI Rates Following Arthroplasty Procedures. IDWeek. San Diego, CA. October 2015.

Calderwood MS, Kleinman K, Bruce CB, et al. Evaluation of Diagnosis Codes Used to Identify Candidate SSI Events during CMS Validation. SHEA Spring Conference. Atlanta, GA. May 2016.

Calderwood MS, Kleinman K, Murphy MV, et al. Not All Colon Procedures Are Equal: Implication for Risk Adjustment in Publically Reported SSI Rates. SHEA Spring Conference. Atlanta, GA. May 2016.

Publications

Letourneau AR, Calderwood MS, Huang SS, et al. Harnessing claims to improve detection of surgical site infections following hysterectomy and colorectal surgery. Infect Control Hosp Epidemiol. 2013;34(12):1321-3.

Calderwood MS, Kleinman K, Murphy MV, et al. Improving public reporting and data validation for complex surgical site infections after coronary artery bypass graft surgery and hip arthroplasty. Open Forum Infect Dis. 2014;1(3):ofu106.

Calderwood MS, Kleinman K, Bratzler DW, et al. Medicare claims can be used to identify U.S. hospitals with higher rates of surgical site infection following vascular surgery. Med Care. 2014. Oct 52(10):918-25.

Calderwood MS, Kleinman K, Huang SS, et al. Surgical Site Infections: Volume-Outcome Relationship and Year-to-Year Stability of Performance Rankings. Med Care. 2017;55(1):79-85.

Calderwood MS, Huang SS, Keller V, et al. Variable case detection and many unreported cases of surgical site infection following colon surgery and abdominal hysterectomy in a statewide validation. Infect Control Hosp Epidemiol. 2017 Sept; 38(9): 1091-1097.

Yokoe DS, Avery TR, Platt R, et al. Ranking Hospitals Based on Colon Surgery and Abdominal Hysterectomy Surgical Site Infection Outcomes: Impact of Limiting Surveillance to the Operative Hospital. Clinical Infectious Diseases 2018 Mar 16, Epub ahead of print.

Calderwood MS, Yokoe DS, Murphy MV, et al. Effectiveness of a Multi-State Quality Improvement Campaign in Reducing Risk of Surgical Site Infections following Hip and Knee Arthroplasty. BMJ Quality & Safety. In revision.