Final Contract Report

Improving Measurement of Surgical Site Infection Risk Stratification/Outcome Detection

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

Final Project Report

Improving the Measurement of Surgical Site Infection Risk Stratification/Outcome Detection

Prepared for:

Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Centers for Disease Control and Prevention
1600 Clifton Rd. Atlanta, GA 30333
www.cdc.gov

Contract No. 290-2006-00-20, Task Order No. 8

Prepared by:

Denver Health
In collaboration with:
Intermountain Healthcare
Salt Lake City VAMC
Vail Valley Medical Center

Principal Investigators:
Connie Savor Price, MD
Lucy A. Savitz, PhD, MBA

AHRQ Publication No. 12-0046-EF
August 2012

Suggested citation:

Price CS, Savitz LA. Improving the Masurement of Surgical Site Infection Risk Stratification/Outcome Detection. Final Report (Prepared by Denver Health and its partners under Contract No. 290-2006-00-20). AHRQ Publication No. 12-0046-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2012.

Contents

Exhibits

Executive Summary

The purpose of this project was to explore opportunities for enhancing the detection and surveillance of inpatient-acquired surgical site infections (SSIs) for four target procedures—herniorrhaphy, coronary artery bypass graft (CABG), and hip and knee arthroplasty (including primary total arthroplasty, primary hemiarthroplasty, and revision procedures). Four delivery systems came together in order to provide the most representative results and generalizable tools. Collaborating delivery systems include Denver Health (a safety-net hospital located in Denver, CO), Intermountain Healthcare (a large, nonprofit, integrated delivery system based in Salt Lake City, UT), and the Salt Lake City Veterans Affairs Medical Center (a VAMC hospital located in Salt Lake City); representativeness was further extended by including the Vail Valley Medical Center (Vail, CO), a Denver Health partner.

One key activity involved developing and testing a computer-assisted algorithm for retrospective assessment of medical records, laboratory test results, and patient demographic data to identify electronically patients with potential SSIs. The plan for, results, and implications of this work is presented in Chapter 2.

A second primary activity of our work involved screening an expanded set of SSI risk factors to consider in exploring, monitoring, and predicting SSIs. The approach for screening additional risk factors, analytic results, and recommendations are presented in Chapter 3.

Chapter 4 provided insights from two target audiences—surgeons and infection prevention nurses—using qualitative research methods to inform Chapter 2 and Chapter 3 activities as well as dissemination and outreach with the results of our work (see Chapter 5).

Chapter 1. Administration

Overall project management and administration was conducted via co-leadership of the Principal Investigators and Project Directors from Denver Health and Intermountain Healthcare, Connie Savor Price, M.D. and Lucy A. Savitz, Ph.D., M.B.A. Denver Health acted as the prime contractor and was responsible for oversight of the collaboration, which was funded through the Accelerating Change and Transformation in Organizations and Networks (ACTION) mechanism. Susan L. Moore, M.S.P.H., served as the ACTION project coordinator and as such was the primary point of contact in facilitating and coordinating communications with the Agency for Healthcare Research and Quality (AHRQ), the Centers for Disease Control and Prevention (CDC), and the research team.

To maintain clear and regular communication, monthly conference calls were held between key project stakeholders, including Kendall Hall, M.D., who represented AHRQ as the Task Order Officer (TOO); Sandra Berrios-Torres, Teresa Horan, and Jonathan Edwards, who together represented CDC as Technical Experts; and the research team leadership from Denver Health, Intermountain Healthcare, Salt Lake City VAMC, and Vail Valley Medical Center. The calls were scheduled on a regular basis to occur on the third Tuesday of each month. A call-in number was established and used for all calls. Agendas and supporting materials for the calls were distributed to the group every Friday before the scheduled calls, to allow sufficient time to review agenda items. Appendix A[*] provides agendas and meeting summaries for these calls.

Following the first teleconference with the TOO and Technical Experts, the project team determined that internal preparation and coordination teleconferences should be scheduled on a regular basis. The internal team calls were scheduled on the second Tuesday of every month. To ensure adherence to the project work plan, the team call agendas were driven by the project work plan timeline and deliverable schedule. The progress of each task was discussed during each call, and any issues or challenges of the work were discussed as a team.

The team-building value of face-to-face meetings was considered important, both early in the project and on a regular basis throughout. The first team meeting was scheduled in Salt Lake City, Utah on November, 20, 2009. At the meeting, the overall project goals were discussed, project management and logistics were reviewed, and the progress of each task was discussed at length. A second meeting was held in Salt Lake City, Utah on October, 15, 2010. The project milestones and deliverables were reviewed with specific focus on the project timeline and end date. Several challenges and updates were discussed at length, most notably, the nursing focus group was proposed to be repurposed to present use cases for the surveillance tool; and the plans for dissemination of the project work were discussed. The TOO and Technical Experts were included in the meeting via teleconference for part of the meeting to review the project updates. A final face-to-face meeting was held in Vail, Colorado on January 27-28, 2011 as a writing session to finalize the draft final report for the project. Summary notes from these in-person meetings are provided in Appendix B.

---

[*] The Appendixes are available on the AHRQ Web site at http://www.ahrq.gov/qual/hais.htm#Providers, as part of the HTML version of the Report. Unlike the Report, the Appendixes are unedited.

Monthly progress reports covered the work done in the previous month and included concise updates of:

- what was accomplished for each task during the reporting period
- problems or delays the contractor has experienced in the conduct of performance requirements, including what specific action is proposed to alleviate the problem(s) and a revised timetable
- activities planned for the next reporting period, including anticipated staffing requirements, level of effort, and cost

To keep all parties informed of the monthly progress, the reports were submitted electronically through AHRQ's Research and Reporting System (ARRS), with duplicates delivered by email to the TOO, Technical Experts, and AHRQ's Contracting Officer. We have explored with AHRQ the opportunity to engage in a variety of dissemination activities. Three presentations, two poster and one oral, have been given at national meetings and two papers are planned for submission to the *Journal of the American Medical Informatics Association* (*JAMIA*) and an open-access, peer-reviewed journal such as *Implementation Science*. This is discussed further in Chapter 5.

Chapter 2. Determine Surgical Site Infection Rates

The purpose of this task is to develop a surveillance tool that detects downstream manifestations of surgical site infection (SSI) in electronic data. Health care systems with electronic health information systems may improve the efficiency (time spent to find a positive case) of their SSI surveillance activities and improve reliability by leveraging electronic data. Although many approaches exist, the one long employed by the Intermountain Healthcare system uses electronic algorithms to populate more manageable queues of charts that an Infection Preventionist (IP) can subsequently review. This approach can capitalize on the IP's superior ability to discern the presence of SSI and may unburden the IP of mundane, automatic tasks significantly. When Intermountain Health initially implemented this scheme, there were few facilities that could have replicated the feat, but more and more facilities may be able to employ similar strategies. Human–adjudicated electronic surveillance for SSI may now be generalizable to other institutions as more hospitals switch over to electronic medical records (EMRs) or electronic health records (EHRs).

The purpose of an IP performing manual surveillance is at least two-fold: to improve situational awareness and to be able to detect differences between times or places. For the former, it is useful to have a high a sensitive surveillance system. To perform the latter, it is useful to be highly specific. IPs have been traditionally employed in this task because they are adaptable and have a better ability to discriminate between charts that have and do not have SSI than automated systems. Although employing IP appears to be the ideal solution, they are routinely stretched and not allowed adequate time for all of their responsibilities.[1-3] Also, adaptation can lead to problems when it comes to comparability.

On the other hand, completely automated systems can review charts quickly and usually do not adapt. There is some evidence to suggest that depending on the conditions and the purpose of surveillance, automated systems may be the instrument of choice.[4] These systems can be extremely sensitive to artifacts of data manipulation or changes in practice. Usually, algorithms are restricted to structured data and cannot use as much information as manual systems. Also, their specificities are usually inferior to manual review.

Our approach used a hybrid, human-adjudicated approach. Such systems are not new, but there have been barriers to implementation. They still need human reviewers, and adapting algorithms developed elsewhere to the local electronic health system may be difficult. The rationale for the combination of the two may be illustrated by invoking signal detection theory.

In signal detection theory, reviewers distinguish between the presence or absence of disease by assessing the chart, laboratory values, antibiotics, etc. These data are called signal. The reviewer has two important characteristics: the discriminability index and criterion. The discriminability index is a measure of how well the reviewer perceives the differences in signal between the diseased and nondiseased states. The criterion is the threshold at which the reviewer interprets signal as disease. If the criterion is lowered, then sensitivity improves and specificity declines. If the criterion is raised then the reverse is true. The only way to improve sensitivity and specificity simultaneously is to improve discriminability. A human reviewer's discriminability index is unlikely to change rapidly, but criterion might. An automated system's discriminability index is usually lower than a human reviewer's, but it can review a large number

of cases rapidly. Its criterion usually does not change unless the semantics of the data have changed. With this framework, we can build a two tiered system. The first tier is run by the automated system. It removes charts where signal is weak enough that they can still be safely removed despite its inferior discriminability index. The second tier involves human review on more difficult cases, where the human's superior discriminability index can be used efficiently.

The algorithm we constructed was developed for use in this two-tiered system. A strength of our approach was to rigorously differentiate between risk factors for and manifestations of SSI. Risk factor data could supply additional information to improve performance, but it would also curtail any analysis of risk from surveillance systems using the algorithm. We anticipated that the main characteristics that would facilitate its acceptability were a high sensitivity and a low number of charts needed to review per identified SSI.

Our approach seeks to capitalize on the superior specificity of human reviewers, the growing wealth of electronic data, and the speed of automated systems. If charts are reviewed in roughly 20 minutes[5] and the fraction of SSI among procedures is roughly 1 percent,[6] then 33 hours of review could be anticipated for every SSI found. If electronic tools could effectively remove 80 percent of charts, then only 6.6 hours would be spent for every SSI found. The impact of such savings may be large. The Virginia requirement for statewide detection/reporting would require 160 infection preventionists (IPs) at a cost of $11.5 million. More than 50 percent of IP time is spent at the desk[7]—time that could be applied to implementation, education, and other effective activities. The surveillance tool will enhance nurse work, moving them from being infection counters to being IPs, freeing these professionals up to do more prevention. Further, the surveillance system provides cognitive surveillance support of the human element of current practice (i.e., chart review, available electronic data, using "shoe leather").

Exhibit 1. Advantages of surveillance automation

- Quality assurance for current practice
- Reduces burden of chart review
- Identify patterns of infection that might suggest opportunities for process improvement/reengineering to enhance quality and safety
- Changes the nature of the job
- Meets mandatory, hospitalwide reporting of SSI for value-based payments;
- Publicly available electronic surveillance tool vs. expensive, proprietary data mining surveillance tools like Theradoc™ or Medimined® that can cost up to $150,000, require a separate server, and have continuing maintenance/upgrade fees

SSI = surgical site infection.

Subtask 2.1. Identify Potential Automated/Electronic Sources of Health Care Data Useful for Surveillance of SSI.

The investigators and supervising officers decided to investigate a human-adjudicated electronic system akin to the one currently used in Intermountain Healthcare. In such systems, electronic algorithms with high sensitivity and negative predictive value are employed to identify electronic markers of SSI and populate a manageable queue of charts that an IP would

subsequently review. The decision to implement this hybrid type of surveillance system was made based on the difficulty of categorizing SSI subtypes and concern for the poor specificity of solely electronic approaches. Initially, the plan was to perform this task using data from the four participating hospitals; however, as we acquired SSI data from these hospitals, it became apparent that the rarity of SSI among the total number of procedures performed would make these data insufficient for the proper training and validation of the electronic algorithm component of our system (see Exhibit 2 below).

In Exhibit 2, it can be seen that there were 73 SSIs, with the smallest hospital contributing only 3 (4.1 percent) and the largest hospital contributing 43 (58.9 percent). The use of data from only the principal four hospitals would produce algorithms based on small numbers and dominated by Intermountain Health.

The standard approach to SSI surveillance, as implemented by the National Surgical Quality Improvement Program (NSQIP), facilitates a consistent measurement of SSI across facilities. Their methodologies are well-documented and there is a certification process for each reviewer. Additionally, NSQIP performs yearly audits to assess interrater reliability. The accrual of data contributed is also reviewed to maintain data quality. The entire Veterans Affairs (VA) network of 152 active hospitals participates in NSQIP (its implementation is called VASQIP) and performs a large number of surgeries. We subsequently received permission for and obtained VASQIP data for 2007 through 2010 for training and testing. Data from the four principal hospitals were used for external validation process described in Task 2.3.2.

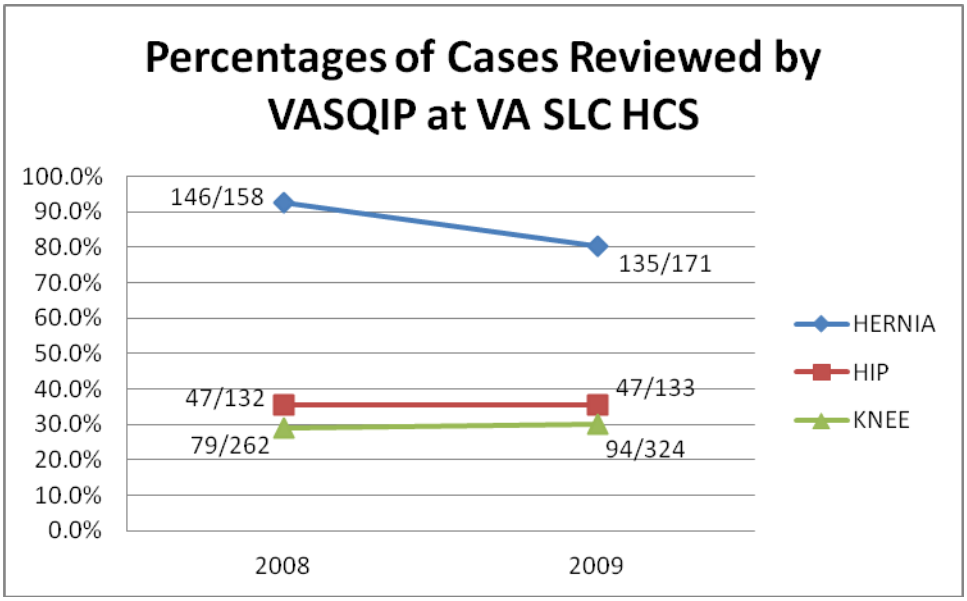
Of note, VASQIP reviewers do not review all cases. Surgical chart reviewers review cases in temporal order as they are identified by CPT code. Reviewing stops when they reach their quotas over 8 day cycles. In the VA, the quota is 36 procedures. No more than 5 of the procedures can be inguinal herniorrhaphies during a cycle.[8] Although sampling is not random, the first day of each cycle shifts the weekday it falls on for each cycle, so it is not obvious that this process produces systematic bias with regard to SSI outcomes. Exhibit 3 illustrates the fraction of cases reviewed among the listed procedures at VA SLC HCS, whether documented only by ICD-9 codes, CPT codes, or both. Depending on the total number of cases of each type, we expect that there would be differences in the proportions reviewed that may vary over time. We cannot exclude bias and it appears that, for some procedures, only a minority of cases is reviewed, but the number of both surgeries and SSIs in hospitals of different sizes, locations, and acuity were seen as an asset for training algorithms to detect SSI. Again, because our objective is to develop an algorithm to detect SSI, the only bias we are concerned with is whether cases are sampled in ways that induce differential misclassification between diagnoses of SSI and our measured indicators of SSI.

Exhibit 2. Number of procedures stratified by hospital and types between 2008 and 2009

Procedure		DH Num.	DH % Proc.	IH Num.	IH % Proc.	VA SLC Num.	VA SLC % Proc.	VVMC* Num.	VVMC* % Proc.
CABG	Total Procedures	0		1845		78		0	
	Superficial SSI	*0*		*12*	*0.7%*	*3*	*3.8%*	0	
	Deep SSI	*0*		*7*	*0.4%*	*0*	*0.0%*	0	
	Organ-Space SSI	*0*		*1*	*0.1%*	*0*	*0.0%*	*0*	
	Total SSI	0		20	1.1%	3	3.8%	*0*	
HERNIA	Total Procedures	898		1059		237		294	
	Superficial SSI	*4*	*0.4%*	*0*	*0.0%*	*2*	*0.8%*	*1*	*0.3%*
	Deep SSI	*2*	*0.2%*	*0*	*0.0%*	*0*	*0.0%*	*0*	*0.0%*
	Organ-Space SSI	*1*	*0.1%*	*0*	*0.0%*	*1*	*0.4%*	*0*	*0.0%*
	Total SSI	7	0.8%	0	0.0%	3	1.3%	1	0.3%
THA	Total Procedures	268		2810		90		137	
	Superficial SSI	*2*	*0.7%*	*0*	*0.0%*	*0*	*0.0%*	*0*	*0.0%*
	Deep SSI	*2*	*0.7%*	*5*	*0.2%*	*0*	*0.0%*	*1*	*0.7%*
	Organ-Space SSI	*3*	*1.1%*	*3*	*0.1%*	*2*	*2.2%*	*0*	*0.0%*
	Total SSI	7	2.6%	8	0.3%	2	2.2%	1	0.7%
TKA	Total Procedures	232		7897		163		421	
	Superficial SSI	*1*	*0.4%*	*6*	*0.1%*	*0*	*0.0%*	*0*	*0.0%*
	Deep SSI	*2*	*0.9%*	*7*	*0.1%*	*0*	*0.0%*	*1*	*0.2%*
	Organ-Space SSI	*1*	*0.4%*	*2*	*0.0%*	*1*	*0.6%*	*0*	*0.0%*
	Total SSI	4	1.7%	15	0.2%	1	0.6%	1	0.2%

DH = Denver Health, IH = Intermountain Health, VA SLC = VA Salt Lake City Healthcare System, VVMC = Vail Valley Medical Center, % Proc = percent of total procedures, CABG = coronary artery bypass grafting, HERNIA = herniorrhaphy, THA = total hip arthroplasty, TKA = total knee arthroplasty.
* From 2007 to 2009.

Exhibit 3. Comparison of fraction of all herniorrhaphy and total knee and hip arthroplasties reviewed by VASQIP



Note: All procedures were identified by ICD-9 and CPT codes. VA SLC HCS = Veterans Administration Salt Lake City Health Care System; VASQIP = Veterans Affairs Surgical Quality Improvement Program

2.1.1. Identify data elements for inclusion in training datasets

A literature review was performed using Medline and the searches " ("surgical wound infection/diagnosis"[Mesh] AND "Data collection"[Mesh])" as well as "("surgical wound infection/diagnosis"[Mesh] AND ("Blood Sedimentation"[Mesh] OR "C-Reactive Protein"[Mesh] OR "Leukocytosis"[Mesh])", which produced 256 and 75 results respectively. Titles and abstracts were reviewed to identify articles to investigate. Our criterion was to identify articles that pertained to the manifestations of surgical site infections, especially those manifestations that can be identified electronically, as opposed to risk factors of disease. We also incorporated articles that the authors were aware of and allowed "snowballing" of related articles during review. We excluded articles that employed primary data collection. The identified data elements were: leukocyte count, leukocyte differential, fever, procalcitonin (not helpful, as this laboratory measurement is not readily available in the United States), erythrocyte sedimentation rate, C-reactive protein, microbiology results, and antimicrobial administration.[9-43] A significant number of articles incorporated claims data into algorithms.[16,18,20,21,26,30,34,44] Unfortunately, these data are generally not available until well after an IP typically would be reviewing cases. We have elected not to include claims data into the algorithm here.

Based on our findings in the literature review, a data dictionary was sent to each of the participating centers, so that they could pull their data. Standardizing to a common physical data model allowed us to share the algorithm through the dissemination of SQL (structured query language) code scripts. Each of the centers then implemented the algorithm on their own data and identified charts that needed to be reviewed. The full data dictionary is included in Appendix C.

2.1.2. Each site pulls surgical-procedural and other identified data, based on the list for training sets

Some modifications were made to this subtask with approval from AHRQ. The movement of large sets of individual data between institutions was problematic. Instead, we focused on developing a portable algorithm, so that each of the centers would be able to implement it locally. Local (as opposed to central) implementation also demonstrates the feasibility of algorithm dissemination.

During the task, it became apparent that VA SLC HCS was the only hospital contributing to NSQIP for all four surgeries of interest. Intermountain Healthcare (IH) does participate in NSQIP, but does not perform NSQIP surveillance on all surgery types of interest. Neither Denver Health (DH) nor Vail participated in NSQIP. At the VA, VASQIP is the principal method of SSI surveillance. For the purposes of training and validating an algorithm, we needed a dataset much larger than the participating hospitals could provide so we decided to use nationwide VASQIP data. A database of vitals, laboratories, medications, microbiology data, and SSI outcomes was constructed for the purposes of algorithm development. Drawbacks of developing an algorithm entirely in the VA and applying it to other hospitals include the fact that the veteran patient population does not necessarily generalize well to the populace at large, and that the VA system has a comprehensive inpatient and outpatient system. IH and DH are similar in this respect. The use of both inpatient and outpatient data improve postdischarge surveillance while making it more efficient. We anticipated that the algorithms might fare more poorly at Vail Valley Medical Center (VVMC) due to the nonintegrated inpatient and outpatient care systems.

A significant amount of work was devoted to ensuring data quality. Since VA data came from many individual hospitals with different data formats, the distributions of each of the data elements was examined to look for outliers. These outliers were then examined to see if they stemmed from differences in units or other unanticipated formats. The final distributions of each were within anticipated bounds.

Subtask 2.2. Develop Procedure-specific Algorithms Utilizing Identified Data Sources to Detect SSI Events

2.2.1. Create and train the algorithm

Our objective was to build an algorithm with high negative predictive value that favored sensitivity over specificity, and relied on human adjudication to improve the specificity of the SSI surveillance system. Traditionally, prospective surveillance systems that rely on manual human review have suffered from suboptimal sensitivity. Because SSIs are rare outcomes, many hours are spent to find each infection, which is extremely inefficient and time consuming. Further, sensitivity may be low because of reviewer fatigue.[45] A human-adjudicated system that reduces workload by removing charts unlikely to contain SSI both reduces the amount of work necessary to detect SSIs and raises the reviewer's expectation that a chart might contain an SSI.

Previous experience detecting methicillin-resistant *Staphylococcus aureus* (MRSA) by means of electronic algorithms[46–48] guided our efforts to find electronic signs of infection as opposed to risk factors. We began with identifying candidate surgeries among VASQIP data from 2007 through 2009. As VASQIP surgeries are identified by CPT code and not by ICD-9s, we built a

map between the two for the four target procedures: coronary artery bypass grafts, total hip arthroplasties, total knee arthroplasties, and abdominal and inguinal herniorrhaphies. We used the UMLS (Unified Medical Language System) metathesaurus concepts to bridge between ICD-9 and CPT vocabularies at the level of coronary artery bypass grafting (CABG), herniorrhaphies, procedures of the hip, and procedures of the distal femur and knee. We then reviewed the children of these concepts and identified codes that described the types of procedures that were included in the ICD-9 list. We felt that this procedure was more reproducible and updatable than an entirely manual mapping attempt. Details of our findings while performing this mapping are included in Appendix D.

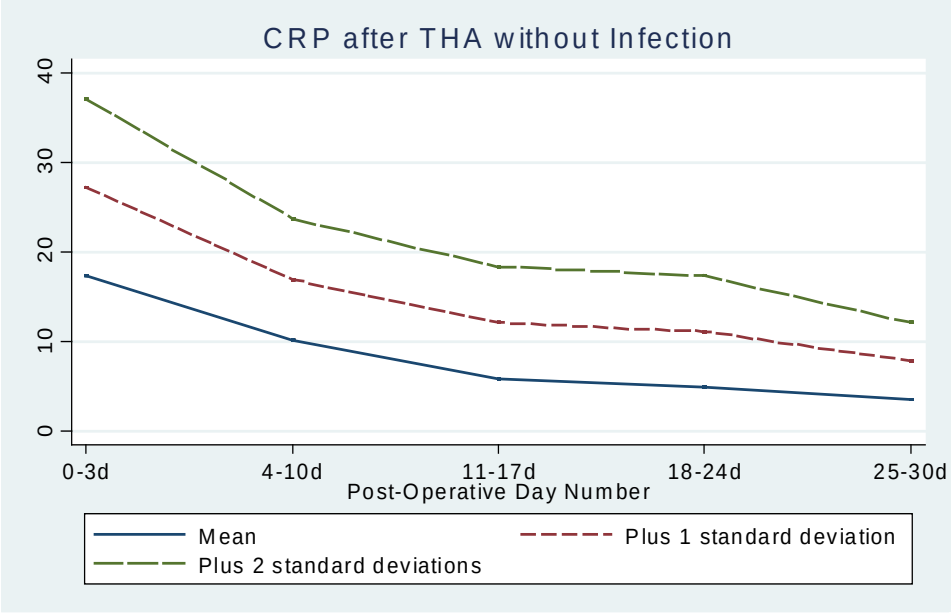
Once the necessary CPT codes were identified, they were used to identify candidate surgeries among all VA hospitals. Between 2007 and 2009, there were 71,102 targeted procedures performed and reviewed in our sampling of the VASQIP dataset. This set was randomly divided into two equal sets, one for training and one for testing. However, due to gaps in the laboratory data we received, the cases before January 1, 2008, were excluded from the testing set.

The dataset also noted whether a superficial, deep, or organ-space SSI was identified within 30 days of the surgical procedure. For simplicity, we summarized the information present in the different levels of infection into dichotomous variables indicating the presence or absence of deep or organ-space SSI, or SSI of any type. As an IP would still need to review the chart, we felt that it was unnecessary for an algorithm to be trained to find each SSI type separately.

These data were then linked to potential manifestations of disease. We included electronic markers between postoperative days 4 and 30 because pre- or perioperative data might indicate risks for SSI or that the patient was already infected at the time of operation. We then investigated the relationship of leukocyte count, temperature, the sending of a microbiology culture, whether the culture matched, the administration of an antibiotic (inpatient or outpatient), readmission, erythrocyte sedimentation rate, and C-reactive protein to SSI. Maximum values during the eligible time-frame were used for laboratory values and vitals. The administration of an antibiotic was limited to systemic antibacterials and readmission was limited to admission to ICU or acute care medical or surgery wards. Although we recognized that risk factors could be associated with SSI as well, we were concerned about introducing mathematical coupling[49]—that bias would be introduced into any subsequent analyses of risk, because risk was used to determine eligibility for SSI.

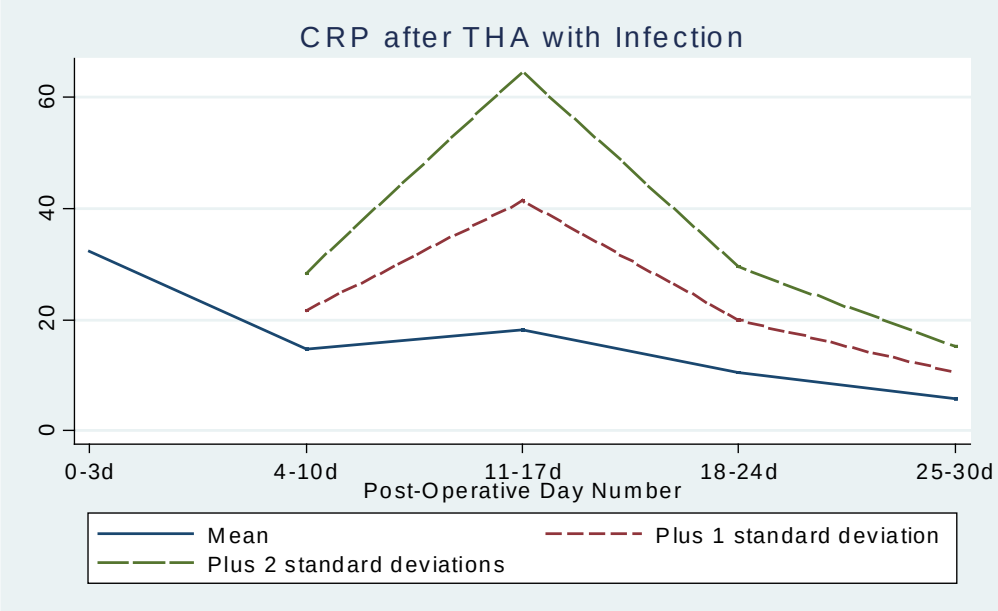
We also considered the potential need for rules that considered the dynamic evolution of the patient's status over time. We excluded the first three postoperative days when determining the administration of antibiotics and sending of cultures because of recommendations for antibiotics prophylaxis and because some operations are performed on known or suspected infected joints. Laboratories and vitals were more difficult because successful operations without complication are known to cause abnormalities that resolve over time. Exhibits 4 and 5 show the evolution of C-reactive protein in total hip arthroplasties both without SSI and with SSI. The laboratory values appeared to have poor correlation with outcome during preliminary analysis, so it was unclear whether this line of analysis would yield much extra information. We opted to simplify by not considering this aspect.

Exhibit 4. Evolution of C-reactive protein values among individuals who NEVER develop an SSI.



CRP = C-reactive protein; SSI = surgical site infection; THA = total hip arthroplasty.

Exhibit 5. Evolution of C-reactive protein among individuals who DO develop an SSI.



CRP = C-reactive protein; SSI = surgical site infection; THA = total hip arthroplasty.

To increase the amount of information that a microbiology culture could provide and to improve the specificity of electronic algorithms, we mapped the reported sample and specimen fields to a single collection-site type. Each type was categorized as to whether it could be consistent with each of the surgeries of interest (see Appendix E). For example, a urine specimen was considered to be incompatible with an SSI from any of our surgeries of interest. A wound swab was considered to be compatible with any of the surgeries. Synovial fluid from the hip was considered to be only compatible with an SSI after a total hip arthroplasty. While all of the

cultures were mapped, not all were considered to be postoperative. Only postoperative cultures were included for consideration in the algorithm. We also extracted information regarding whether there was growth of any organism, whether there was growth of a virulent organism, and whether the specimen came from a normally sterile site. But, as it became clear that implementation would be difficult at other centers, we discontinued development of this aspect.

Various strategies are available for algorithm development. We targeted algorithms with high sensitivity that also could increase the efficiency of chart review by excluding a large fraction of negative charts. To do the latter while not impeding the former, we investigated methods that would allow interactions between variables. Classification tree and regression tree (CART) analysis, also called recursive partitioning, lends itself to the formulation of interacting rules and has been used previously in algorithms to detect SSI.[16] This method is limited, however, because it does not analyze interactions along the entire range of variables. Another issue is that it is not as robust when dealing with frequent missing data. Random forest strategies may have had advantages, but we felt that, for user acceptability, it was important to have simple, understandable rules.

We used the function rpart for recursive partitioning in R, to develop algorithms. We used the function initially to detect all types of SSI, but because of the lack of sensitivity and inefficiencies when searching for superficial SSIs, it was subsequently trained to target only deep and organ-space SSIs. It was felt that the reliability in the reference standard would also be higher in this subgroup.[50] The results presented in this document refer to these later algorithms; however, classification trees of these earlier attempts are also included in Appendix F. We specified a classification tree and a loss matrix to penalize false negatives. The loss matrix was weighted by the inverse of the prevalence of deep and organ-space infections in the set. The maximum depth was limited to three and the minimum number of cases in a branch, before a split was permitted, was three. Any tree that resulted in a change of the complexity parameter (cp) of more than 0.001 was investigated. Effort was taken to prune the tree at the cp that minimized the relative cross-validation error, but when the difference was small and the algorithm was not sensitive enough, values with more splits but slightly higher relative cross-validation errors were accepted. The model was built with the following R code:

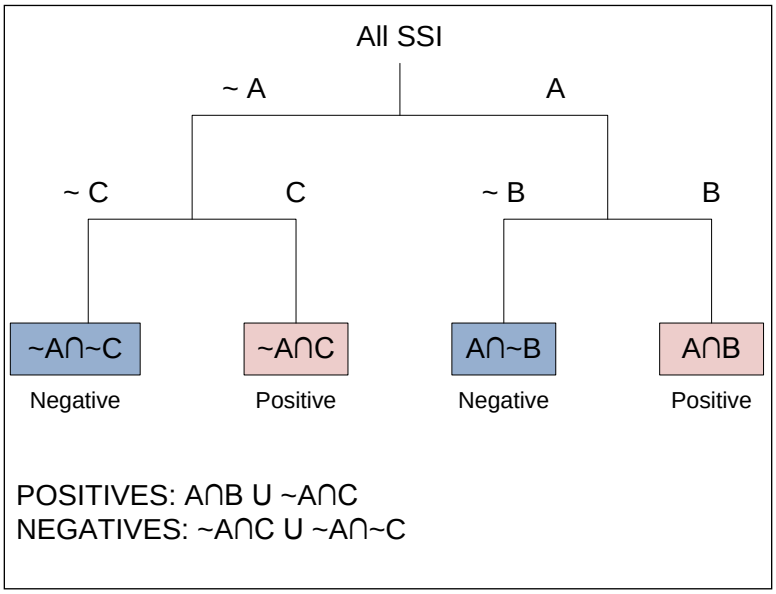
```
inverse_prevalence<-(1/mean(N$SSI))
loss<-matrix(c(0,inverse_prevalence,1,0),byrow=TRUE,ncol=2)

fit<-rpart(SSI~WBC +ESR +CRP +NE_N +NE_P  +FERRITIN
+P_WBC +P_ESR +P_CRP +P_FERRITIN +postopabx +postopcx +postopadmit
,data=N, minsplit=3, maxdepth=3, method="class", cp=.001, parms=list(loss=loss))
```

Where ‘N’ is the original data frame and ‘N$SSI’ is the vector representing deep and organ-space SSI. ‘loss’ is the loss matrix. The modification of the variable names with the prefix ‘P_’ indicates the presence of a value, as we anticipated that the presence or absence of a lab may be informative as well. NE_N denotes the absolute number of neutrophils and NE_P denotes the percentage of neutrophils.

The initial tree, the correlation coefficient and cross-validation errors, and the final tree (if different from the initial tree) are included in Appendices F & G. The presentation of rules in Exhibit 6 is equivalent to the charts in the appendix. This is because the tree has been collapsed into an expression of the set of surgeries among which SSI are likely.

Exhibit 6. Demonstration that algorithm positives can be identified by compact logic.



A, B, and C represent conditions. ~ is the NOT operator. ∩ is the INTERSECTION operator and U is the UNION operator. SSI = surgical site infection.

The way in which classification trees are collapsed is illustrated in Exhibit 6, if the first branch includes on A and its subbranch includes on B, then the set A∩B meets both conditions, where ∩ is the INTERSECTION operator. Because of the law of commutation, A∩B= B∩A and, by extension, A∩B∩C= C∩B∩A. Each branch that extends to the right joins the new set by the INTERSECTION operator. The rest of the figure describes how rules may be expanded. Exhibit 7 presents the same algorithm as the charts in the appendix, except that the law of commutation has been used to make the rules more presentable. An important point to make is that missing values do not evaluate to true or to false. They evaluate to NULL. All of these NULLs are also interpreted as positive.

Exhibit 7. Algorithm to identify deep and organ-space surgical site infections

Surgery	Description
CABG	1. Presence of both postoperative culture and postoperative antibiotics, and the maximum postoperative leukocyte count is not less than 11.85
Herniorrhaphy	One of the following: 1. Presence of a postoperative culture and one of the following criteria: a. The maximum postoperative leukocyte count is not less than 7.78 b. The maximum postoperative neutrophil percentage is not less than 67.3 2. Absence of a postoperative culture and one of the following criteria: a. Postoperative antibiotics given and any postoperative leukocyte count test drawn b. Postoperative antibiotics not given, but the patient had a postoperative admission
TKA	One of the following: 1. Presence of a postoperative culture 2. Presence of a C-reactive protein and the maximum postoperative leukocyte count is not less than 9.45
THA	1. Presence of a postoperative culture and postoperative antibiotics were given and the maximum postoperative leukocyte count is not less than 7.55

CABG = coronary artery bypass graft; TKA = total knee arthroplasty; THA = total hip arthroplasty.

Alternatively, we tried an "inclusive" rule using the presence of any high-normal value.

"Inclusive" rule:

- the presence of an erythrocyte sedimentation rate (ESR) greater than 20
- or a total neutrophil count greater than 5,000/mm^3
- or a leukocyte counter greater than 9,000/mm^3
- or a C-reactive protein greater than 3mg/dL
- or postoperative antibiotics given
- or the presence of a postoperative culture, or the patient was readmitted within 30 days postoperatively.

Finally, we also implemented a "simple" rule.

"Simple" rule:

- Microbiology test ordered between postoperative days 4 and 30, inclusively
- An antibacterial was prescribed between postoperative days 4 and 30, inclusively

The algorithms' performance on the training set can be seen in Exhibits 8–10 below. In Exhibit 8, the total numbers of procedures are divided by procedure type. Additionally, the breakdown by depth of infection and fraction of total SSI are expressed. In Exhibit 9, two-by-two tables and diagnostic accuracy in the training set are listed. Additionally, the fraction excluded and the numbers of charts that need to be reviewed per positive case are expressed for both unfiltered review and algorithm filtered review. The sensitivity was as low as 93.3 percent for herniorrhaphies and total knee arthroplasties. When heavily penalizing false-negatives, even beyond the inverse prevalence, the rpart function would not return acceptable algorithms. The "inclusive algorithm" was implemented as well, which is detailed above, but did not employ any

other logic than logical OR/set union. Exhibit 10 demonstrates the gains in sensitivity from a very inclusive algorithm. While there were some gains in sensitivity, particularly with respect to superficial SSI, which we were not targeting, this came at the expense of needing to review approximately one-third of the charts.

Exhibit 8. Distribution of SSI among surgeries in VASQIP data

	CABG		HERNIA		KNEE		HIP		TOTAL	
Total Procedures	4525		19692		7467		3867		35551	
Surgical Site Infections	# SSI	% Total	# SSI	% Total	# SSI	% Total	# SSI	% Total	# SSI	% Total
sSSI	108	93.1	165	74.7	61	59.2	40	51.3	374	72.2
dSSI	8	6.9	36	16.3	29	28.2	33	42.3	106	20.5
oSSI	0	0.0	20	9.0	13	12.6	5	6.4	38	7.3
TOTAL	116		221		103		78		518	

SSI = surgical site infections, VASQIP = Veterans Affairs-Surgical Quality Improvement Program, CABG = coronary artery bypass grafting, HERNIA = herniorraphy, KNEE = total knee arthroplasty, HIP = total hip arthroplasty, #SSI = number of SSI, % Total = percent of all SSI, sSSI = superficial SSI, dSSI = deep SSI, oSSI = organ space SSI

Exhibit 9. Preliminary performance on VASQIP training set

	CABG		**HERNIA**		**KNEE**		**HIP**	
Total Cases	4525		19692		7467		3867	
Positive SSI	TP 8	FN 0	TP 52	FN 4	TP 39	FN 3	TP 36	FN 2
Negative SSI	FP 402	TN 4115	FP 1004	TN 18632	FP 873	TN 6552	FP 295	TN 3534
ID by Alg	410	4115	1056	18636	912	6555	331	3536
%Excluded by Alg	90.94		94.64		87.79		91.44	
Diagnostic Accuracy								
Sensitivity	100.00%		92.86%		92.86%		94.74%	
Specificity	91.10%		94.89%		88.24%		92.30%	
PPV	1.95%		4.92%		4.28%		10.88%	
NPV	100.00%		99.98%		99.95%		99.94%	
AUROC	95.60%		93.90%		90.50%		93.50%	

	CABG	HERNIA	KNEE	HIP
Time Saving				
NNR w/Alg	51.25	20.31	23.38	9.19
NNR s/Alg	555.56	357.14	178.57	102.04

Alg = algorithm, AUROC = area under receiver-operator curve, CABG = coronary artery bypass grafting, HERNIA = herniorraphy, ID = identified, KNEE = total knee arthroplasty, HIP = total hip arthroplasty, NNR= number needed to read, NPV = negative predictive value, PPV = positive predictive value.

Exhibit 10. Performance of different algorithms by SSI type on the training set

	Total Surgeries	sSSI	dSSI	oSSI
Total	35551	374	106	38
	Surgeries Flagged	# Caught (Sensitivity)		
Rpart Algorithm	2709 (7.6%)	156 (41.7%)	101 (95.3%)	34 (89.5%)
Inclusive Algorithm	7961 (22.4%)	328 (87.7%)	104 (98.1%)	35 (92.1%)
Simple	6939 (19.5%)	318 (85.0%)	103 (97.2%)	34 (89.5%)

SSI = surgical site infection. sSSI = superficial SSI. dSSI = deep SSI. oSSI = organ-space SSI. Inclusive algorithm looks for any high-normal value. Rpart algorithm refers to the algorithm derived using the Rpart procedure. ( ) = percentage of total

After the algorithm was developed, its sensitivity was 93.8% (95% confidence interval [CI] = 88.5–97.1) and its specificity was 93.0% (95% CI = 92.7–93.3) for all procedures compared to the training NSQIP dataset. Its positive and negative predictive values were 5.2 (95% CI = 4.3–6.1) and 99.99% (95% CI = 99.9–100), respectively. Thus, when an IP reviews these procedures, we would expect her/him to review 18.9 charts on average before finding an SSI using the recursive partitioning algorithm, 79.3 using the "inclusive" algorithm, and 246.9 if all charts were reviewed.

2.2.2. Externally validate data elements for inclusion in the algorithm & create final list

The VA NSQIP data was randomly divided into two equal size sets for validation, because a second set of data was not collected prospectively. Data from the VA Salt Lake City Healthcare System were excluded because they would later be used in the analysis of our four principal hospitals of interest. We had initially considered bootstrap validation. However, when we decided on using NSQIP data and anticipated a much larger number of outcomes, we elected to use one-fold cross-validation.

Sensitivity, specificity, and positive and negative predictive value were calculated by comparing the electronic algorithm's output against the testing set to report final validation numbers. Its sensitivity and specificity were 73.1 percent and 92.9 percent, respectively. The

positive predictive value (PPV) was 3.9 percent and the negative predictive value (NPV) was 99.9 percent. Unfortunately, the statistics for sensitivity were well below those seen on the training set. However, the inclusive and simple algorithms' performance remained stable, as seen when comparing Exhibits 10 and 11.

Exhibit 11. Performance on SSI types using test set

	Total Surgeries	sSSI	dSSI	oSSI
Total	27,739	279	75	33
	Surgeries Flagged	# Caught (Sensitivity)		
Rpart Algorithm	2034 (7.3%)	126 (45.2%)	54 (72.0%)	25 (75.8%)
Inclusive Algorithm	8,685 (31.3%)	251 (90.0%)	75 (100%)	30 (90.9%)
Simple	5318 (19.2%)	230 (82.4%)	68 (90.7%)	30 (90.9%)

SSI = Surgical Site Infection. sSSI = superficial SSI. dSSI = deep SSI. oSSI = organ-space SSI. Inclusive algorithm looks for any high-normal value. rpart algorithm refers to the algorithm derived using the rpart procedure. ( ) = percent of total

Not all elements were included in the final algorithm. Although we initially planned to include fever in the algorithm, Denver Health did not have this information extending back through the whole cohort so it was removed from analysis. To develop an easily comprehensible algorithm with face validity and easier implementation, we targeted an algorithm with a minimal set of easily pulled elements (see Exhibit 12).

Exhibit 12. Data for algorithm

Name	Short Description	Units/variable type	Long Description
Hosp_ID	Hospital identifier	varchar	Unique identifier for the hospital
Pt_ID	Patient identifier	varchar	Unique identifier for each patient
SGY_ID	Surgery identifier	varchar	Unique identifier for each surgery
sSSI	superficial surgical site infection	integer [0,1]	0 if routine surveillance found no infection; 1 for + superficial SSI
dSSI	deep surgical site infection	integer [0,1]	0 if routine surveillance found no infection; 1 for + deep SSI
oSSI	organ space surgical site infection	integer [0,1]	0 if routine surveillance found no infection; 1 for + organ space SSI
postopabx	postoperative antibiotics	integer [0,1]	1 if any antibiotics given between postop day 3 and 30, but not day -1 to -7; else 0
postopcx	postoperative culture	integer [0,1]	1 if any microbiology culture sent between postop day 3 and 30; else 0
postopadmit	postoperative readmission	integer [0,1]	1 if any readmission between postop day 3 and 30; else 0
wbc	white blood cell count	float [K/mm3]	maximum recorded wbc between postop day 4 and 30

Name	Short Description	Units/variable type	Long Description
crp	C-reactive protein	float [mg/dL]	maximum recorded crp between postop day 4 and 30
SGY_Type	Procedure Type	varchar	[CABG, THA, TKA, Herniorrhaphy]

Each of the hospitals was sent the data elements necessary for the final algorithm. Actual code scripts were also sent to facilitate algorithm implementation; however, we realized, as others have before,[51, 52] that tailoring and adjustments would have to be made to accommodate different data structures at each facility. The final algorithm was implemented in the SQL script included in Appendices F and G.

2.2.3. Create test database from final/validated list of data elements

As previously mentioned, it became apparent that the movement of individual level data between institutions for this task was an unworkable option. In lieu of this, we focused on algorithm portability. Instead of a single database upon which to run the electronic algorithm, multiple local databases and local implementation of electronic algorithms solved the issue of moving individual-level data back and forth. Each institution performed data pulls for the necessary elements. Scripts that encoded algorithms for each of the four target procedures were written and distributed to each of the facilities. Data and computer professionals at each hospital tailored the code to run on their data. Each center's code was reviewed by the team at Salt Lake VAMC in order to catch potential misunderstandings during the implementation and adaptation process. Finally, the electronic algorithm was used on local hospital data, which identified charts to be reviewed for Subtask 2.3.

Subtask 2.3. Demonstrate Performance of Electronic Detection Schemes Compared to an Accepted SSI Surveillance Reference Standard.

2.3.1. Acquire all surveillance data pertaining to cases in the test set

Each of the facilities had different preexisting strategies for SSI surveillance. Denver Health and Vail hospitals followed National Healthcare Safety Network (NHSN) guidelines and performed traditional, manual surveillance. Intermountain Healthcare previously pioneered electronically supported, human-adjudicated surveillance systems and used that modality routinely across all hospitals in their system. The VA used NSQIP for surveillance with rules similar to, but not entirely the same as NHSN. None of the participating hospitals followed NHSN guidelines pertaining to not counting procedures where a drain was placed through the incision intraoperatively. Each of the facilities pulled the results of routine surveillance based on their own methodologies into databases residing on their own systems. Each of these datasets served as slightly different reference standards representing the status quo. As such, accuracy and reliability statistics between the centers are not directly comparable, but represent the performance of the human-adjudicated surveillance systems compared to the various systems already in place.

With regard to amending VASQIP data to be more in line with NHSN guidelines, it became clear that Denver Health, because of its integrated system (where outpatient and inpatient records

are part of the same EHR) also principally focused on postdischarge surveillance of up to 30 days. Because of variances from and differences in interpretation of the NHSN guidelines at each center, it was deemed difficult to attempt to build or amend data from the routine surveillance data to achieve harmonization to a common standard at each facility. In the end, no consistent reference standard could be applied across all hospitals.

2.3.2. Analysis and reporting of electronic surveillance & manual surveillance performance to the reference standard

At each hospital, we applied our electronic algorithm to all surgical procedures that met our prespecified criteria. Results were pooled to report overall accuracy. The sensitivity was 37.8 percent, the specificity was 94.3 percent, the PPV was 2.0 percent, and the NPV was 99.8 percent. Results from each system calculated individually are outlined in Exhibit 13. The sensitivity ranged from 0 percent at VVMC to 50 percent at VA SLC HCS. As the numbers of positives are quite small, the confidence intervals for sensitivity and PPV are quite large.

Exhibit 13. Accuracy of algorithm at each participating hospital

Accuracy of Algorithm at DH

		Routine Surveillance		
		SSI	no SSI	Total
Algorithm	SSI	6	71	77
	no SSI	7	1345	1352
	Total	13	1416	1429

Sensitivity	46.2%
Specificity	95.0%
Positive Predictive Value	7.8%
Negative Predictive Value	99.5%

Accuracy of Algorithm at IH

		Routine Surveillance		
		SSI	no SSI	Total
Algorithm	SSI	9	704	713
	no SSI	16	10857	10873
	Total	25	11561	11586

Sensitivity	36.0%
Specificity	93.7%
Positive Predictive Value	1.3%
Negative Predictive Value	99.9%

Accuracy of Algorithm at VA SLCHCS

		Routine Surveillance		
		SSI	no SSI	Total
Algorithm	SSI	2	33	35
	no SSI	2	531	533
	Total	4	564	568

Sensitivity	50.0%
Specificity	94.1%
Positive Predictive Value	5.7%
Negative Predictive Value	99.6%

Accuracy of Algorithm at VVMC

		Routine Surveillance		
		SSI	no SSI	Total
Algorithm	SSI	0	17	17
	no SSI	3	832	835
	Total	3	849	852

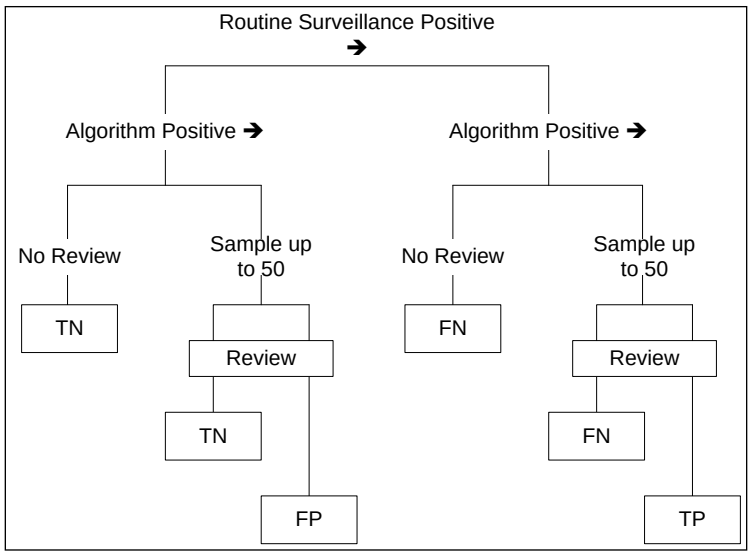
Sensitivity	0.0%
Specificity	98.0%
Positive Predictive Value	0.0%
Negative Predictive Value	99.6%

DH = Denver Health; IH = Intermountain Healthcare; VA SLC HCS = Veterans Affairs Salt Lake City Health Care System; VVMC = Vail Valley Medical Center; SSI = surgical site infection.

We compared the human-adjudicated system to routine prospective surveillance as a test of feasibility and diagnostic performance. Of all of the charts that the algorithm deemed positive and queued for review, we randomly selected up to 50 that had also been identified by routine surveillance as positive (*i.e.*, true positives) and up to 50 identified as negative (*i.e.*, false positives) for manual review at each center. The reviewer was blinded as to the result of routine surveillance as well as to the ratio of positives and negatives. The reviewer classified each chart as to whether an SSI was present and the depth of the SSI. The procedure type was already specified by ICD-9 or CPT code. Charts not queued for review by the algorithm were considered negative by the human-adjudicated system. Exhibit 14 illustrates the sampling and reviewing processes. Sensitivity, specificity, and their confidence intervals were calculated using methods included in the Appendix H. We found a sensitivity of 41.5 percent (excluding records with corrupted identifiers), which is lower than hoped for, but had been limited only by the insensitivity of the electronic algorithm. Our specificity of 99.8 percent was comparable to that frequently reported by manual surveillance systems. The overall measured interrater reliability between the two historical surveillance assessments and the assessments of our reviewers on sampled charts was 0.85.

During implementation, poor algorithm performance was noted at VVMC (Exhibit 15). At most other facilities, the absence of an antibiotic prescription after surgery meant either no antibiotic was given, or that the prescription data were missing from the record. In the case of VVMC, no electronic antibiotic prescription data were available, so all were missing. In their case, we coded '-1's in the postoperative antibiotic field and altered the algorithm to allow '-1's to cause the algorithm to err on the side of calling cases positive for flagging. No cases were picked up with the changes.

Exhibit 14. Schematic showing the sampling process for chart review and the assignment of true and false positives and negatives*



* Relative to routine surveillance, as the reference standard.

TP = true positive;, TN = true negative; FP = false positive; FN = false negative.

Exhibit 15. Accuracy of adjudicated algorithm at each participating hospital

Accuracy of Adjudication at DH

		Routine Surveillance		
		SSI	no SSI	Total
Adjudicated	SSI	6	1.42	7.42
	no SSI	3	1410	1348
	Total	9	1411.4	1420.4

Sensitivity	66.7%
Specificity	99.9%
Positive Predictive Value	80.9%
Negative Predictive Value	99.8%

Accuracy of Adjudication at IH

		Routine Surveillance		
		SSI	no SSI	Total
Adjudicated	SSI	9	0	713
	no SSI	16	11559	11575
	Total	25	11559	11584

Sensitivity	36.0%
Specificity	100.0%
Positive Predictive Value	100.0%
Negative Predictive Value	99.9%

Accuracy of Adjudication at VA SLCHCS

		Routine Surveillance		
		SSI	no SSI	Total
Adjudicated	SSI	2	4	35
	no SSI	2	560	562
	Total	4	564	568

Sensitivity	50.0%
Specificity	99.3%
Positive Predictive Value	33.3%
Negative Predictive Value	99.6%

Accuracy of Adjudication at VVMC

	Routine Surveillance
Adjudicated	No Change

DH = Denver Health; IH = Intermountain Healthcare; VA SLC HCS = Veterans Affairs Salt Lake City Health Care System; VVMC = Vail Valley Medical Center; SSI = surgical site infection.

Exhibit 16. The rpart algorithm alone—sensitivity by surgical procedure type

	DH			IH			VA SLC			VVMC		
	TP	FN	SN	TP	FN	SN	TP	FN	SN	TP	FN	SN
CABG	NA	NA	NA	4	4	50.0%	0	0	NA	0	0	NA
HERNIA	3	3	50.0%	0	0	NA	1	0	100.0%	0	1	0.0%
THA	1	3	25.0%	3	5	37.5%	0	2	0.0%	0	1	0.0%
TKA	2	1	66.7%	2	7	22.2%	1	0	100.0%	0	1	0.0%

DH = Denver Health, IH = Intermountain Health, VA SLC = VA Salt Lake City Healthcare System, VVMC = Vail Valley Medical Center, CABG = coronary artery bypass grafting, HERNIA = herniorrhaphy, THA = total hip arthroplasty, TKA = total hip arthoplasty, NA = not applicable, TP = true positive, FN = false negative, SN = sensitivity

Exhibit 17. Sensitivity and positive predictive values of algorithms alone

	IH		VA SLC		DH	
Algorithm	SN	PPV	SN	PPV	SN	PPV
rpart	10/25 (40%)	10/896 (1.1%)	2/4 (50%)	2/35 (5.7%)	6/13 (46.2%)	6/77 (7.8%)
Simple	13/25 (52%)	13/1100 (1.2%)	2/4 (50%)	2/106 (1.9%)	5/13 (38.5%)	5/157 (3.2%)
Inclusive	21/25 (84%)	21/7250 (0.3%)	2/4 (50%)	2/184 (1.1%)	11/13 (84.6%)	11/381 (2.9%)

IH = Intermountain Health. VA SLC = VA Salt Lake City Health Care System. DH = Denver Health. rpart = algorithm developed using recursive partitioning. SN = sensitivity. PPV = positive predictive value.

All reports of sensitivity, especially when grouped by procedure as in Exhibit 16, must be interpreted with extreme caution. The confidence intervals are wide. For example, at VVMC the algorithm found 0 of 3 SSI. These data would be observed with a sensitivity of up to 60 percent over 5 percent of the time.

All positives identified by the algorithm, as well as positives identified by routine surveillance were reviewed (Exhibit 17). At DH, the study reviewer agreed with all of the cases identified as positive by routine surveillance. Four surgeries were noted to have incorrect ICD-9 codes, indicating that they should not have been included. The study reviewer also identified one superficial SSI and one deep SSI caught by the algorithm, but not found in routine surveillance records. At VA SLC HCS, four additional deep and organ-space SSI were identified in addition to those identified by routine surveillance. At VVMC, all algorithm-identified cases were false positives. At IH, the study reviewer agreed with all positive cases identified by routine surveillance and none identified by algorithm, except for two cases where there appeared to have been errors with identifiers.

False negatives were reviewed at each center to determine the reasons for failure and to identify areas for future algorithm improvement. At Denver Health, two of the false negatives represented problems with the data pull. One SSI was assigned to the wrong hip replacement in the historical dataset. The hip replacement with infection was not in the dataset. Another procedure identified as having an SSI was actually a hysterectomy. Three surgeries were missed because the SSI occurred greater than 30 days postoperatively. One SSI was missed because laboratories were only available from the outpatient setting. One SSI could only have been picked up from emergency department notes. Only two SSI could have been picked up by electronic data, but were missed due to the algorithm's threshold criteria.

At VA SLC HCS, only two SSI were missed. Both occurred in total hip arthroplasties with onset of infection greater than 30 days postoperatively. This is interesting because although VASQIP protocol is to extend surveillance through 30 days, it appears that SSIs occurring outside of this time period are being recorded. Fortunately, algorithm sensitivity can be easily increased in this setting by increasing the observation period.

At VVMC, the algorithm was unable to detect three SSIs, due to one surgery being treated solely in the outpatient setting and another being treated at an outside facility. The last infection developed 11 months after surgery and was thus not picked up because of the time period of surveillance.

At IH, 11/16 false negatives were attributed to the algorithm missing important information. Most of this information was in the clinical notes (10/12), with the remainder in microbiology. All of this information occurred after discharge from the initial surgery. In 2 cases, the reviewer felt that the cases were ambiguous, in another 2 the reviewer actually disagreed that the cases were SSI. In 1 case, the reviewer felt that the case actually was sSSI rather than dSSI or oSSI.

The single most remediable element for the algorithm is the duration of surveillance after surgery. Otherwise, it appears that further improvements may be difficult without information extraction from free-text notes. Difficulties that arise from fragmented care are not easily addressed, but do contribute to missed SSIs.

Subtask 2.4. Assessability of Electronic Detection Methods to Determine Procedure-specific, Organism-specific SSI Rates Versus Estimates of Annual National Burden and Identifying SSI in Health Care Facilities

2.4.1. Identifying SSI in health care facilities

We tabulated identified organisms by procedure at each hospital. Organisms associated with deep- or organ-space SSI were counted by positive microbiology cultures from postoperative day 4 to 30. If more than one organism was present, all were counted. No attempt at establishing the organism as the etiology of the SSI was made. Organisms were grouped in Exhibit 18. Trend analysis was not feasible because of the short-time period represented in the data and the rarity of events, especially when divided by procedure.

Exhibit 18. Counts of identified organisms by procedure*

Surgery	Organism	DH	IH†	VA SLC	VVMC
CABG	*S. aureus*	NA	1	0	NA
	Enterococcus	NA	1	0	NA
	P. aeruginosa	NA	0	0	NA
	Other	NA	5	1	NA
HERNIA	*S. aureus*	1	0	1	0
	Enterococcus	0	0	0	0
	P. aeruginosa	0	0	0	0
	Other	2	3	1	0
TKA	*S. aureus*	1	3	2	0
	Enterococcus	0	1	0	0
	P. aeruginosa	0	0	0	0
	Other	0	2	0	1
THA	*S. aureus*	3	3	0	1

Surgery	Organism	DH	IH†	VA SLC	VVMC
	Enterococcus	0	2	0	0
	P. aeruginosa	0	1	0	0
	Other	1	2	0	0
SSI with recognized pathogens		7	20	4	2
Total SSI		13	25	4	3

DH = Denver Health; IH = Intermountain Health; VA SLC = Veterans Affairs Salt Lake City Health Care System; VVMC = Vail Valley Medical Center; CABG = coronary artery bypass grafting, HERNIA = herniorrhaphy, THA = total hip arthroplasty, TKA = total hip arthoplasty,

*More than one organism can be counted for a single SSI. Only identified organisms are counted. †Organisms categorized by surgical bed, so that other surgeries with SSI at the same site could not be easily differentiated.

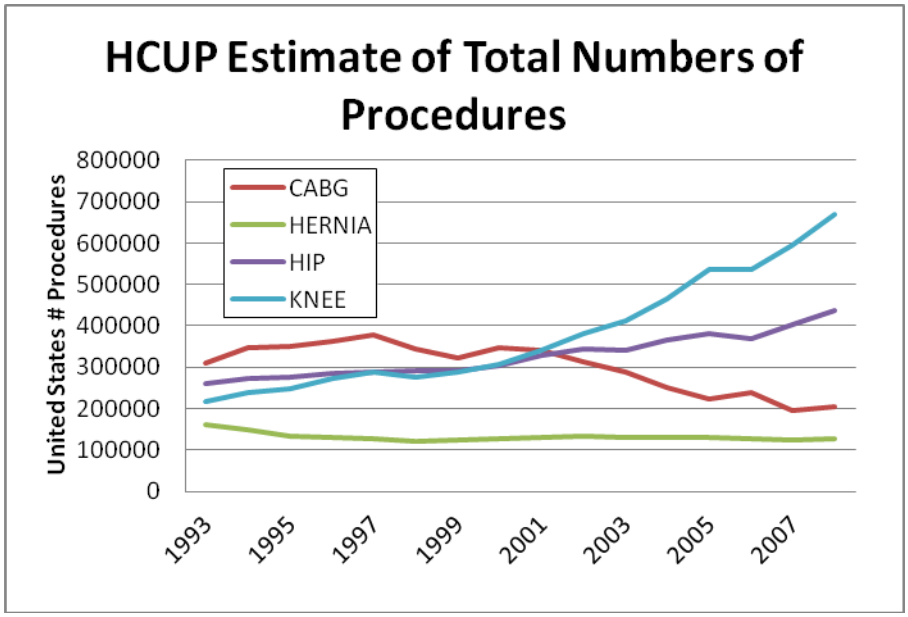
2.4.2. Estimate the national procedure-specific, organism-specific rates of SSI

We used methods similar to those outlined by the National Nosocomial Infection Surveillance System (NNIS) as a model for estimating the annual national burden of SSI for the specified four procedures.[53] The SSI incidence proportions are listed in Exhibit 20. We did differentiate between ICU and non-ICU infections as in the NNIS study because of the difficulty with assigning hospitalization days that were at risk for SSI versus days that were a result of SSI. Inpatient sample data from AHRQ's Healthcare Cost and Utilization Project (HCUP) were used to estimate the number of procedures performed nationwide in 2008 (Exhibit 19). Multiplying the number of procedures by the pooled incidence proportion gave us 16,530 as an estimate of nationwide yearly instances of SSI for these four procedures. Individually, they were 4,478 for CABG, 1,233 for herniorrhaphies, 4,929 for total knee arthroplasties, and 5,890 for total hip arthroplasties. These numbers are not directly comparable to NNIS results because we are only examining a fraction of the surgeries. Additionally, the four hospitals were selected for convenience and do not represent the range of variability in hospitals across the country (by size, case mix, etc.). Exhibit 21 contrasts the number of total knee arthroplasties performed at each of the principal hospitals with the average number of procedures performed at various hospital types using HCUP data. Deficiencies in representativeness are evident. Combining these into a single estimate of SSI rate is misleading, at best and could be quite biased. Accurate estimates could be made by defining segments in the population thought to have different SSI rates (this is a challenge in and of itself) and sampling hospitals from each. At the very least, to apply this method to determine national SSI burden, there would need to be a larger number of hospitals chosen to represent the variability in size and case mix. Reporting confidence intervals would be inappropriate given these limitations.

There are severe limitations to this approach and there should be caution with interpretation. Unfortunately, the nationwide SSI rates are unknown–the best estimate being one from 2002 by NNIS.[53] Our smaller group of hospitals is only informative in a very limited way. Vail was noted to have the lowest risk patients and a very low SSI rate. If all of the hospitals had the same SSI rate, then a low estimate of infections associated with 254 herniorraphies, with 2,620 THAs, and with 1,337 TKAs could be made, but not without a large amount of uncertainty secondary to very small numbers. Estimates using each of the hospital's rates and pooled rates are illustrated

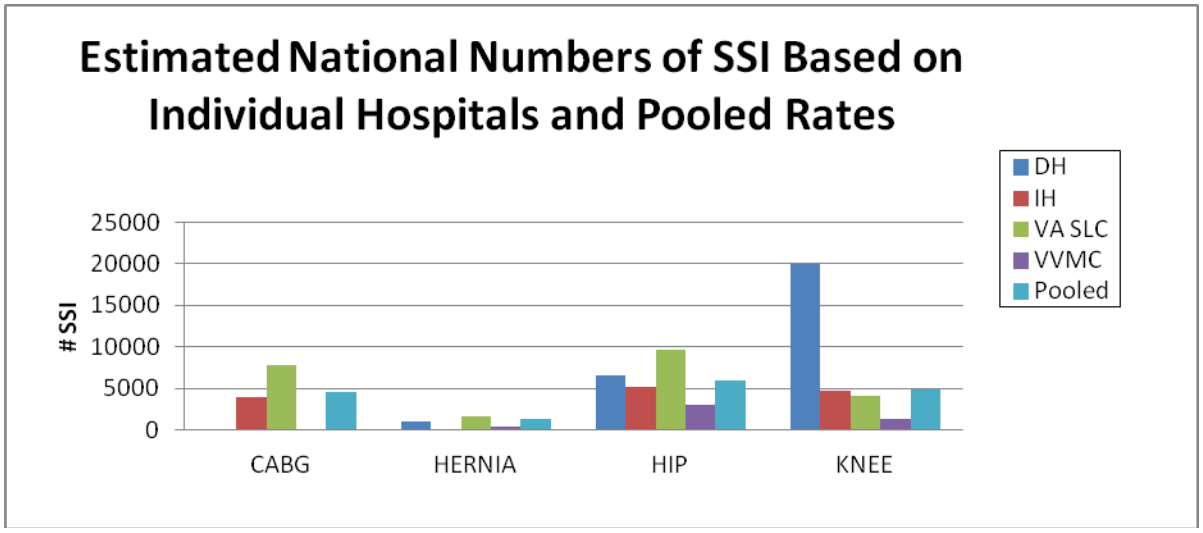
in Exhibit 19 and Exhibit 20 below. Again, more accurate estimates would require sophisticated patient case mix adjustment and a much larger sampling of hospitals.

Exhibit 19. HCUP estimates of the total number of target procedures



CABG = coronary artery bypass grafting, HERNIA = herniorrhaphy; HIP = total hip arthroplasty; KNEE =total knee arthroplasty.

Exhibit 20. Estimates of the total number of SSI cases in the United States, based on the rates used.



DH = Denver Health; IH = Intermountain Health; VA SLC = VA Salt Lake City Health Care System; VVMC = Vail Valley Medical Center. Pooled ; pooled rate over all four hospitals; SSI = surgical site infection(s).

Exhibit 21. No. of total knee arthroplasties at principal hospitals in 2008 compared to the HCUP estimates for different hospital types

Rank of 2008 #TKA Discharges	1	2	3	4	5
Hospital Type					
# Hospitals in Hospital Type	256	308	1301	627	23
% Nation's TKA at this Hospital Type	15.4%	14.3%	29.4%	14.2%	NA
# TKA Discharges 2008	513.6	394.8	192.4	192.4	191.3*
Rank of 2008 #TKA Discharges	6	7	8	9	10
# Hospitals in Hospital Type	1	220	1	187	1072
% Nation's TKA at this Hospital Type	NA	3.6%	NA	2.4%	11.4%
# TKA Discharges 2008	140.3*	139.3	116*	109.9	90.5
Rank of 2008 #TKA Discharges	11	12	13		
# Hospitals in Hospital Type	1	969	724		
% Nation's TKA at this Hospital Type	NA	7.7%	1.7%		
# TKA Discharges 2008	81.5*	67.8	19.6		

* Averaged between 2008 and 2009. Highlighted columns represent hospitals in our dataset.
TKA = total knee arthroplasties, VA = Department of Veterans Affairs.

The results above are methodologically problematic, but are best estimates considering the data and resource limitations of the task. Sound estimates would necessitate a large enough dataset to estimate the rate at each hospital with confidence. Ideally (for statistics, but not for

patients), there would be enough events, so that the rate could be estimated with confidence with at least one year granularity. Facilities that are too small to accumulate enough events to estimate their rates within an aggregated 2- to 3-year time period are difficult to use when secular trends are important. The sample of hospitals needs to be large enough to provide power in the estimation of rates and be representative of the target population. Some adjustments can be made if patient-case mix is considered. However, the number of SSIs at each hospital is small and we had only four hospitals from the Intermountain west to analyze. Acquiring the data for and developing valid models that adjust for patient-case mix is a large endeavor outside of the scope of this task. It is perhaps more appropriate to compare the individual SSI rates that we found with that reported in the literature.

Total hip arthroplasty. In much of the developed world, the incidence of SSI after total hip arthroplasty appears low. Reports from the United States range from 0.9 to 2.52 percent.[54,55] In Europe, rates have been reported as low as 1.3 percent in Germany[24] and as high as 3.4 percent in the Netherlands.[56] In Brazil, one small study reported a rate of 17 percent for this procedure.[57] Unfortunately, with this and other studies, there may be some important differences in reporting systems that complicate the direct comparisons of rates from one institution to the next. Also, the point estimates from small single-hospital studies may be spurious and not reflect the true underlying rate because of chance. However, there does appear to be real variability between institutions and countries. The rates we reported from each of our hospitals range from below the low end reported in the United States to that reported in the Netherlands.

Total knee arthroplasty. Similar trends were seen in TKA as in THA. The rates for SSI after TKA range from 0.5 percent in France[58] to 0.99–2.3 in the Netherlands.[56] Using Medicare data, the rate of prosthetic joint infections was estimated at 1.55 percent within 2 years of surgery.[59] As with THA, TKA had a higher rate in Brazil from the same study above.[57] As with THA, we observed rates that ranged from below the low to above the high range reported in these first world countries. This is not surprising, given the small numbers in this study compared to the studies cited.

Coronary artery bypass grafting. Reports of anywhere between 0.560 and 2.7 percent can be seen for deep SSI after cardiac surgery. A study in New York found a rate of 1.8 percent after isolated CABG and 2.4 percent after CABG combined with other cardiac procedures, such as valve replacement.[61] Only two centers performed CABG in our study. Our values of 1.9 and 4.2 percent are comparable to, but perhaps a bit higher than, that reported in the literature. One aspect to point out is that Intermountain Health rates are calculated from data already derived from an algorithm-driven, human-adjudicated process, which may decrease review fatigue and improve sensitivity. As with the other surgeries, direct comparisons can be difficult because followup time can be variable.

Herniorrhaphy. A Cochrane review in 2007 identified four studies that used CDC criteria when considering antimicrobial prophylaxis for herniorrhaphy or hernioplasty. The other studies used various criteria that were not completely in line with CDC guidelines. Among the four studies, the incidence proportion of SSI among those receiving prophylaxis was 0–1.72 percent and 1.9–8.16 percent, respectively, for herniorrhaphy and hernioplasty. The reported proportions from our four hospitals are closest to those reported in the literature for patients receiving preoperative antibiotics.

Subtask 2.5. Estimate Burdens of Electronic SSI Detection Relative to an Accepted Reference Standard

Our ability to conduct a detailed cost estimate was limited, in that costs associated with information technology implementation are largely system-specific and are dependent on available resources and the expertise of inhouse personnel. As a result, we present a general categorization of the expected costs and resources required for adaptation and implementation of an electronic SSI detection tool, and provide an estimate of cost savings due to successful implementation for an example case.

As previously mentioned, we observed variations in SSI surveillance practices across the four partner health delivery systems. The practice of conducting manual review for surveillance with no automated support was used as the standard base case. Cost and resource use for implementation of the electronic surveillance algorithm were categorized in two primary cost buckets: (1) cost to set up; and (2) costs to run. In addition, we also note the expectation that there will be future costs and a resource burden associated with regular maintenance and required updating of the algorithm, to maintain the value and accuracy of the tool.

The SSI-identifying algorithm is a program written in structured query language (SQL) that uses Boolean logic (in "case" statements) to identify potential surgical site infections. Setting up the query is a one time, up-front investment that results in a generalized program that can be reused repeatedly. The time required to set up the algorithm varies, based on the electronic availability of required data fields within a given delivery system, as well as the expertise of personnel available to perform the algorithm implementation. If all the variables are easily accessible and experienced personnel available, the algorithm can be programmed in 1–2 hours. The most difficult aspect of developing the query is the complexity inherent in joining multiple data sources such as laboratory, surgery, and patient demographic data.

DH serves as a case example for adapting and implementing an electronic surveillance algorithm in a health care delivery system in which manual review is used as standard surveillance practice.

DH identified 2,179 procedures for inclusion in its system-specific cohort. Sixty procedures were associated with SSI after manual chart review using NHSN methodology. The modified algorithm flagged 804 procedures (or 37 percent of total charts) for review. The percent of total procedures flagged for review varied by procedure type, and ranged from 15 percent (herniorrhaphy) to 62 percent (craniotomy). The modified algorithm achieved 100 percent sensitivity and 72 percent specificity in detecting SSIs, validated on 4 years of our manual SSI surveillance data using NHSN methodology.

Based on these results, 1,375 unnecessary chart reviews would have been avoided over the 4-year period without sacrificing detection of a single SSI. Assuming 20 minutes per chart for manual review, 57 work days (8 hours/day) of chart review would have been eliminated using the algorithm for surveillance of SSI in hip and knee arthroplasty, abdominal and vaginal hysterectomy, spinal fusion, craniotomy, and herniorrhaphy.

Task 2 Discussion

Measurements of algorithm diagnostic accuracy must be carefully considered. Routine, manual, prospective surveillance is estimated to have a sensitivity between 30 percent and in excess of 90 percent, with most estimates in the 70–80 percent range.[21,35,50,62,63] Additionally, the reliability of manual health care-associated and surgical site surveillance has been reported to be less than ideal.[21,64-67] For many of these studies, either routine surveillance or some augmentation of routine surveillance is used as the reference standard. Any comparisons to such standards must take this into account. Other electronic algorithms are frequently reported to have sensitivities in excess of 80 percent.[38,68] Only some of them have been applied to multiple hospitals,[16,21,35] and none of them report individual hospital validation results among hospitals as heterogeneous as the principal hospitals of our study. Although our recursive partitioning algorithm had high sensitivity on the VASQIP training set, its sensitivity was 73.1 percent on the VASQIP test set. The pooled sensitivity at the four principal hospitals was 41.5 percent. These results contrast with the high performance seen in other published literature. Specificities and predictive values were relatively stable between training and testing sets.

The differences in sensitivities that we see in the recursive partitioning algorithm suggest two levels of overfitting. The first is overfitting to the training dataset and the other is overfitting to the VA system. The two "common sense" algorithms demonstrated high sensitivity in both the VASQIP training and testing sets. Since they were not derived from the training set, they are not overfit to it. We expected the sensitivity of these algorithms to be high because of success for previously devised algorithms—and because we surmised that it was unlikely that patients with deep or organ-space SSI would neither be treated by antimicrobials nor tested for etiologic microorganisms. However, when these algorithms were tested against other hospitals, the sensitivity and positive predictive values varied. At VA SLC HCS, no improvement in sensitivity was observed over the recursive partitioning algorithm, perhaps due to small numbers. At IH, a much larger number of false positives was found, resulting in a much lower positive predictive value. This appears largely due to a different pattern of antimicrobial use during the postoperative period in this system. At DH, the simple algorithm fared poorly, while the inclusive algorithm fared better. This illustrates that the "common-sense" algorithms that include elements successful at other institutions[16,21,35] still did not generalize well because of institutional differences.

Other factors may also have contributed to the less than desirable performance of the algorithm derived from recursive partitioning. Poor reliability in the reference standard, too little information, and limitations in the recursive partitioning method may also have been factors. Having relatively few SSIs to work with was also a limiting factor for algorithm training.

The amount and quality of electronic information, as well as algorithm development issues, are also important. The available data elements may not have been informative enough. In 1992, Harvard Pilgrim Health used the same recursive partitioning method on a set of 4,086 procedures containing 96 postdischarge SSIs to develop algorithms to predict SSI.[16] The approach allows for some interactions of variables in rule generation. The group allowed variables such as diagnosis codes, age, sex, length of surgery, microbiology, antibiotics, readmissions, and emergency department visits. The principal difference between our variables and the Harvard study was that we omitted diagnosis codes, age, and sex, and included laboratory values. We made these omissions for reasons stated previously. They developed multiple models: one with a sensitivity

of 74 percent and a specificity of 98 percent, one with a sensitivity of 92 percent and a specificity of 92 percent, and another with a sensitivity of 77 percent and a specificity of 94 percent. Each of these models was created by adding different sets of information. The authors noted that no model surpassed a sensitivity of 80 percent without both outpatient diagnosis codes and microbiology results. It is still unclear how well those algorithms would perform on an external set. The discriminability indices of the models above [discussed earlier, on pp. 5–6] were as follows: 2.7, 2.8, and 2.3 respectively. In comparison, our algorithm had a sensitivity of around 73 percent and a specificity of around 92 percent. The "simple" algorithm, using only data on postoperative cultures or postoperative antibiotics, had a sensitivity of 95 percent and a specificity of 81 percent on the same set. The discriminability indices were 2.0 and 2.2, respectively. Similar discriminability indices would suggest that the criterion is drastically different to produce very different sensitivities. The effect of criterion is illustrated by examining the inverse of the positive predictive value, or the number of charts needed to review to find an SSI. IP would likely choose the more sensitive algorithm, although they would need to review an average of 53 charts for every SSI, instead of 23. However, any option would like be better than the approximately 250 charts needed to be reviewed to find a deep or organ-space SSI. Problems with identifying SSI may have been in part because: (1) the available information was insufficient; (2) recursive partitioning may not be robust to missing data and/or not fully consider interactions between variables; and (3) the penalty for false negatives was not high enough.

The algorithm may have problems with generalizability. However the measures of sensitivity at the principal hospitals are difficult to interpret because the confidence intervals are very wide. Improving the algorithm's sensitivity, while keeping the number of charts needed to review low, can only be accomplished by improving the algorithm's discriminability. This could be accomplished by using procedures more robust to sparse data for algorithm development, incorporating dynamic thresholds for laboratory values and vitals, and enriching the input data by using natural language processing to extract information from text notes. Recently, natural language processing was used to find postoperative complications (but not SSI) using VA data.[69]

Another potential approach to improve acceptability is to allow IPs to change criterion by allowing them to choose from a menu of algorithms. Instead of using algorithms to rule out cases by forcing a dichotomy, algorithms would predict the likelihood of SSI to facilitate subsequent triage and chart review. Algorithms could also facilitate review by annotating important aspects of the case.

In this system, IP would need to accept that greater sensitivity comes at the price of reviewing more charts. When discriminability is held constant and criterion is lowered, then sensitivity improves and specificity worsens. If criterion is raised then the opposite happens. Just as we demonstrated with the "inclusive" algorithm, we could investigate other algorithms with known characteristics. If IP are unsatisfied with the sensitivity of an algorithm, then the next most sensitive algorithm can be selected. If IP feel that they are reviewing too many cases, then a more specific algorithm can be selected. The drawback to this approach is that it compromises the reliability that automated systems could contribute.

A modification of this last approach could use strategies similar to IBM's DeepQA technology to incorporate information from multiple, simultaneously applied algorithms for prediction. This technology was used in the recently televised Watson experiment on Jeopardy.[70] Model-combining[71] or multiple-model-estimation[72] approaches could potentially provide better performance, but they would increase the complexity of the rules and make the automated

system a black box. This may be less acceptable to IP, but it could still be used in a decision-support style system. For example, instead of a two-tiered system, a multitiered system could be employed. As before, IP would be the final tier, but the automated system could have multiple tiers. The initial tier would use a high-sensitivity algorithm to completely eliminate a modest number of charts, perhaps up to half. The next tier would flag possible SSI, but not remove charts. One or multiple algorithms could be used in this tier. Finally, IP would still need to review all charts that from passed the first tier, but could choose to spend more or less time on charts, depending on the flags.

The most appropriate use of automated systems, whether alone or in combination with manual surveillance, will take careful consideration of the purpose and requirements of the events being surveilled. The performance of automated systems may vary, particularly when attempting to detect events that occur in the outpatient setting, where differences in data availability may be pronounced. More work is necessary to improve the discriminability index of electronic algorithms, but allowing IP to select rules that suit their own needs may be a reasonable measure in the interim.

Chapter 3. Designing and Testing Methods to Stratify the Risk of Surgical Site Infections

The primary purpose of this task was to design and test methods to risk-stratify surgical patients for surgical site infection. We began by clarifying conditions and data availability for analyses. Exhibit 22, below, summarizes data availability by facility/system. It can be seen that data were available for all target procedures across partner facilities except CABG data that were not available for Denver Health or Vail Valley Medical Center.

Exhibit 22. Availability of data by procedure and facility/system

Procedure	Denver Health	Intermountain	SLC VAMC	*Vail*
CABG	NA	√	√	NA
Herniorrhaphies	√	√	√	√
Hip prosthesis	√	√	√	√
Knee Prosthesis	√	√	√	√

CABG = cardiac artery bypass grafting, NA = not applicable, SLC VAMC = Salt Lake City Veterans Affairs Medical Center

Subtask 3.1. Identify Strong Predictors of SSI, Particularly Important Variables Not Currently Used in Mainstream Risk-stratification Methods.

The list of potential risk factors for surgical site infections was developed with a two-tier process:

An active surgeon on the study team used his extensive experience with SSIs and a previous list of risk factors used by his institution to identify potential risk factors. That initial list consisted of 88 risk factors for SSI (see Appendix I). This list was then used in a focus group to solicit input with surgeons (see Chapter 4).

An extensive literature review was also performed using Internet search engines (including PubMed and Google Scholar) to identify any published risk factors for SSIs at any site. All English-language publications for the previous 10 years were included. Keywords used for the search included: *SSI, surgical site infection, surgical risk factor, risk factor, surgical wound, surgical infection*. Risk factors identified from any surgical site were included in the list. From that search, 24 additional potential risk factors were included in the Master Risk Factor table (see Appendix J). Each of the potential risk factors in the list was then clinically reviewed and categorized as modifiable or nonmodifiable.

Subtask 3.2. Develop a Risk-adjustment Method that Utilizes the Identified Risk Variables to Validly Compare Rates of SSI Across Facilities

The Master List of all identified potential risk factors (see Appendix K) was sent to each of the four study sites—Intermountain Healthcare, Denver Health, Salt Lake VA, and Vail Valley Medical Center. Each site examined the Master List and determined if it had electronic access to each of the individual risk factors. Each site then returned its marked list to Intermountain Healthcare where their site-specific information was added to the Master List. After that information was collected from all four sites, a union set of 34 potential risk factors was identified. From that union set, a new list of risk factors common to all four sites was created (see Appendix L; also, Appendix M compares the initial and final lists of risk factors). Each of those risk factors was then further defined to remove any ambiguity between the study sites, ensuring identical collection and reporting. The data values for each risk factor, their description, and type were agreed upon via conference calls and email. Based on that process, a final data collection spreadsheet was created in Microsoft Excel® and sent to each of the four study sites. Each site then met with their data access colleagues to plan how the data would be collected, and made sure both groups had the same definitions of the needed risk factors. Each site developed programs based on their specific data-retrieval needs to access and collect the data elements needed for each risk factor.

Each site then collected their data for patients older than 18 years of age, using unique patient identifiers for patients who had CABG, herniorrhaphies, hip arthroplasty, and knee arthroplasty. Intermountain also collected data for appendectomies and added that data to their final dataset.

Each patient in the study group from each of the four sites was then identified as having or not having an SSI, based on the specific surgical procedure, and marked accordingly in the spreadsheet. SSI data was collected based on the reporting site's specific collection method, with SSIs defined using NNIS criteria from CDC. A random sample of patients and identifiers were manually selected to verify that data access was correct at each site. Each site then deidentified the data and submitted their final spreadsheet to Intermountain Healthcare. Mappings to actual patient identifiers were kept behind the firewalls at each of the study sites.

All four sites then examined each of the risk factors they had electronic access to and documented the original data source for each. Exhibit 23 denotes data sources for the 33 risk factors at each of the sites. As is often the case during the actual data collection process, some of the risk factors contained in the common list were found not to be stored (or only stored occasionally) in the databases at the different sites. For example, at Intermountain, the American Society of Anesthesiologists ASA score is a data element in the surgery database, but we found it was only populated 17 percent of the time. This was often the case at the other sites also. Thus, ASA score was not included in the final list of 33 common risk factors used in the analysis (see Appendix L). While each of the four sites had access to the common list of risk factors, each site often collected that data from different clinical departments or databases. This information should help other facilities to determine where they may find these risk factors at their institutions. At Intermountain, eight additional risk factors electronically found in the database and not included in the common list were also included in a second spreadsheet (Exhibit 24,

below). Other sites also identified some unique risk factors not found at the other participating sites, but that were not included in this study.

Exhibit 23. Data sources for the SSI common risk factors

Column Heading	Denver Health	SLC VA	Vail	Intermountain
Admit_src	Utilization	PTFMove	Encounter.Admit_src_cd	Demographics
Icu_admit	Utilization	PTFMove	Encntr_loc_hist.nurse_unit_cd	ADT data
Inpat_surg	Utilization	Local VASQIP	Encounter.Encntr_type_cd	Surgery data
Outpat_surg	Utilization	Local VASQIP	Encounter.Encntr_type_cd	Surgery data
Age	Utilization	demographics	Person.birth_dt_tm (calculation)	Demographics
BMI	Pharmacy/ Lab	vitals	Clinical_event	Height & weight
CA_hx	Utilization/ Surgery	Dx Codes	Diagnosis	ICD-9 codes
Chronic kidney dx	Utilization	Dx Codes	Diagnosis	ICD-9 codes
Chronic lung dx	Utilization	Dx Codes	Diagnosis	ICD-9 codes
COPD_hx	Utilization/ Surgery	Dx Codes	Diagnosis	ICD-9 codes
Diabetes	Utilization	Dx Codes	Diagnosis	ICD-9 codes
DVT	Utilization/ Surgery	Dx Codes	Diagnosis	Vascular studies
Hypocholest	Utilization	Dx Codes	Diagnosis	ICD-9 codes
MRSA	Utilization	Microbiology	person_patient.disease_alert_cd	Microbiology
Patid	ADT	PTF	Person.Person_id	Demographics
Payer	Utilization	Fee Basis	Health_plan.health_plan_cd	Demographics
Postop_hematocrit	Lab	laboratory	Clinical_event	Laboratory data
Postop_admit	Utilization/ Surgery	ADT		Laboratory data
Preop_hematocrit	Lab	laboratory	Clinical_event	Laboratory data
Preop_hemoglob	Lab	laboratory	Clinical_event	Laboratory data
Postop_hemoglob	Lab	laboratory	Clinical_event	Laboratory data
Preop_stay	Utilization/surg	ADT	Encounter.arrive_dt_tm Surgical_case.surg_start_dt_tm (calculation)	ADT & surgery
Proeop_albumin	Lab	laboratory	Clinical_event	Laboratory data
Rheum-dx	Utilization	Dx codes	Diagnosis	ICD-9 codes
Sex	Utilization	Demographics	Person.sex_cd	Demographics
SSI	NHSN reports	Local VASQIP	Manual tracking	Infection control
Abx_dc		Pharmacy	Clinical_event	Pharmacy data
ASA	NHSN reports	Local VASQIP	Surgical_case.asa_class_cd	Surgery data
Emergent	NHSN reports/ Surgery	Local VASQIP	Surgical_case.sched_type_cd	Surgery data
General_anes	NHSN reports/ Surgery	Local VASQIP	Surgical_case.anesth_type_cd	Surgery data
No_procedures	Utilization	Local VASQIP	Count instances of surg_case_procedure for a surgical_case	Surgery data
Surg_date/time		Local VASQIP	Surgical_case.surg_start_dt_tm	Surgery data
Surg_dur	Surgery	Local VASQIP	Surg_case_procedure.proc_dur_min	Surgery data
Surg_proc	Utilization	Local VASQIP	Hardcoded based on dx	Surgery data
Surgeon_ex	Medical Records	Local VASQIP	Not done	Surgery data
Wnd_class	Surgery	Local VASQIP	Surgical_case.wound_class_cd	Surgery data

Exhibit 24. List of additional SSI risk factors at Intermountain Healthcare

Column Heading	Value Type	Description
Charlson Score	Numeric	Charlson Score
Preop glucose	Numeric	Preop glucose
No. of People in OR	Numeric	Number of people in OR
Postop glucose	Numeric	First glucose level after surgery
Postop fluid	Numeric	Liters of Fluid 24 hrs PostOp
Abx allergy	Numeric	1 if Hx of abx allergy…0 if not
Trach	Numeric	1 if patient had tracheotomy…0 if not
No. of surgeons	Numeric	No. of surgeons scrubbed

Abx = antibiotics; Hx = history; OR = operating room; Postop = postoperative; Preop = preoperative

Description of Intermountain data. All of the data from Intermountain Healthcare was collected from data contained in the Enterprise Data Warehouse (EDW), which resides on an Oracle® relational database. All clinical data from the EMR, surgery database, hospital-acquired infection database and other databases contained in the Intermountain EMR are loaded into the EDW each night. The EDW contains 35,000,000,000 records and 8 Terabytes of data. Each of the common risk factors was collected using SQL queries on specific tables or the union of multiple tables. All Intermountain patients have an enterprise[-wide identification] number that is consistent for all encounters at any of the 22 hospitals and over 100 InstaCare facilities, clinics, and physician offices. That number was used to link all patient specific data. Surgery data was queried first to identify all patients undergoing the study procedures during the study period. Each of those patients was then checked for SSIs linked by the date of the surgery. All the other data elements listed in the common data list (and the eight other data elements available to Intermountain only) were then collected, based on the definitions included in the common list. After the data was then checked and verified, the final study database was loaded and sent for statistical analysis.

Description of Denver Health data. Data were available in the warehouse. The many ancillary services housed in this single database include lab, radiology, pharmacy, scheduling, and surgery. Along with the Web Portal, information in the data warehouse can be accessed through Crystal Reports, Executive View, and Microsoft Analysis Services, as well as by using other tools. These data were collected using Statistical Analysis Software® (SAS®) Enterprise guide, version 4.2. DH data were limited to total knee (TK) and total hip (TH) replacements, and herniorrhaphies (HE), as CABGs are not performed at this location. These data were found in utilization, lab, surgery, and pharmacy electronic repositories. Other variables were found in NHSN reports. Denver Health has a data warehouse with a unique patient identifier and unique episode identifiers used to link data across systems. There are some limitations to DH data. Surgeon experience is limited to the number of years practicing at DH. There are missing datapoints, including 616 for ASA and 427 for surgeon's experience. We were unable to locate data on antibiotics discontinued within 24 hours. There were three out-of-range values found for preop stay length (less than 0 days) and six out-of-range values for surgery duration (less than 0 minutes). Up to 15 diagnosis codes were available per surgery, and up to 10 procedure codes. The limitation on procedure codes is not a large problem, though, as only seven surgeries (0.5 percent) had 10 procedure codes.

Description of Vail data. Data are available via an EMR system (CERNER). This data is accessed via either chart review or through prebuilt reports that require a coded program. The system allows unlimited diagnosis and procedure codes for each surgery. Data was collected using Excel®. VVMC data were limited to total knee (TK) and total hip (TH) replacements and herniorrhaphies (HE), as CABGs are not performed at this institution. These data were found in demographics, ADT, Surgery, Nursing, ICD-9, Microbiology, Laboratory, and Pharmacy modules of our EMR.

There were two limitations to the VVMC data: first, surgeon experience was unable to be collected; and second, the algorithm was run without including postdischarge antibiotic prescription data for a large majority of patients who return home for postdischarge care.

Description of SLC VAMC data. The VA Salt Lake City Health Care System Data Mart is a compilation of operational data designed to extend the utilization of the clinical and administrative systems. The data mart is comprised of a collection of databases storing data from The Veterans Health Information Systems and Technology Architecture (VistA) and other data sources. VistA is an enterprise-wide information system built around an EHR. Data are stored in a relational database. Targeted patient population was selected using both ICD-9 codes and CPT codes (total knee [TK], and total hip [TH] replacements, herniorrhaphies HE], and CABG). Multiple VA data sources are merged and cohorts are definable by attributes such as ICD-9 codes and CPT codes from both inpatient and outpatient encounters within a target time period. These data are kept current by frequent updates with new data from the source databases, so that timely data are available for research. Additionally, surgical outcome data were obtained from VASQIP. VASQIP data represents an extensive surgical quality improvement program and data collection tool. Comprehensive selections of approximately 69 clinical variables are collected for each case in this option. The dataset contains a broad range of variables that can be used for research purposes, as well as for identifying opportunities for surgical process improvement and other quality improvement efforts.

Data collection summary:

A total of 3,612 herniorrhaphies, 3,410 total hip and 9,728 total knee procedures were included in the study using Intermountain, Vail Valley Medical Center, VAMC, and Denver Health. An additional 1,802 CABG and 5,873 appendectomy procedures were submitted from Intermountain and the VAMC (Exhibit 25). A total of 222 SSIs were associated with the various surgical procedures and participating facilities (Exhibit 26). The SSI rates varied by reporting site and procedure each year, and ranged from 0.0 to 7.1 percent.

Exhibit 25. Number of procedures per data site and year

	Surgery Procedure					
	AP	CA	HE	TH	TK	Total
2007						
DH			296	71	76	443
IH					2	2
VA		4	16	3	11	34
Vail			109	42	87	238
2008						
DH			328	88	72	488
IH	2941	922	1057	1345	4210	10475
VA		42	119	45	83	289
Vail			130	47	149	326
2009						
DH			274	109	84	467
IH	2932	802	1002	1548	4587	10871
VA		32	102	42	69	245
Vail			179	70	298	547
Total						
DH			898	268	232	1398
IH	5873	1724	2059	2893	8799	21348
VA		78	237	90	163	568
Vail			418	159	534	1111
All Sites	5873	1802	3612	34109	9728	24425

DH =Denver Health; IH =Intermountain Healthcare; VA =Salt Lake City VA Vail =Vail Valley Medical Center; AP =Appendectomy; CA =CABG; HE =Herniorrhaphy; TH =Total Hip Arthroplasty; TK =Total Knee Arthroplasty

Exhibit 26. Number of SSIs per procedure and data site

	Surgery Procedure					
	AP	CA	HE	TH	TK	Total
2007						
DH	.	.	1	2	3	6
IH	.	.	.	.	0	0
VA	.	0	0	0	0	0
Vail	.	.	0	1	0	1
2008						
DH	.	.	3	0	3	6
IH	14	17	22	19	36	108
VA	.	3	2	2	0	7
Vail	.	.	0	0	0	0
2009						
DH	.	.	3	2	1	6
IH	15	15	7	16	29	82
VA	.	0	1	0	1	2
Vail	.	.	2	0	2	4
	Surgery Procedure					
	AP	**CA**	**HE**	**TH**	**TK**	**Total**
Total						
DH	.	.	7	4	7	18
IH	29	32	29	35	65	190
VA	.	3	3	2	1	9
Vail	.	.	2	1	2	5
All Sites	29	35	41	42	75	222

DH = Denver Health; IH = Intermountain Healthcare; VA = Salt Lake City VA Vail = Vail Valley Medical Center; AP = Appendectomy; CA = CABG; HE =Herniorrhaphy; TH = Total Hip Arthroplasty; TK = Total Knee Arthroplasty

Subtask 3.3. Evaluate the Quality of the Risk-Adjustment Relative to Method Complexity and Data Collection Costs

We met with a statistician and selected appropriate statistical tests to identify the risk factors for SSI. The dataset was checked by the statistician for any obvious coding problems or other issues that would complicate or confound the analysis. The statistician also verified that the dataset was formatted correctly so that it could be loaded into the statistical software for analyses. After the data was loaded into the statistical program, it was then further cleaned and any missing or incomplete data was resolved.

After combining the data from each health care system, we found numerous missing lab values (Exhibit 27), especially for the outpatient procedures. We found that the variables preop_hematocrit, preop_hemoglob, preop_albumin, postop_hemoglob, and postop_heatocrit were most often missing because those tests were not ordered prior to the patient's procedure (Exhibit 28). An analysis of lab values 30 days prior to surgery showed that 99.7 percent of lab values were captured within 30 days preop, and verified that the missing values were indeed the result of the tests not being ordered prior to the procedure. However, dropping an entire record due to missing values can result in undesired outcomes and misleading results,[73,74] therefore we decided to impute the missing values. We considered a bootstrapping approach, but ultimately chose to use a multiple imputation (MI) method to approximate missing data.

Exhibit 27. Summary of missing data

Value	Non-SSI Patients	SSI Patients
abx_dc	1970 (8%)	27 (11%)
admit_src	643 (3%)	9 (4%)
age	642 (3%)	13 (6%)
asa	18954 (76%)	170 (72%)
bmi	4320 (17%)	27 (11%)
ca_hx*	0 (0%)	0 (0%)
ch_kidney_dx*	0 (0%)	0 (0%)
ch_lung_dx*	0 (0%)	0 (0%)
copd_hx*	0 (0%)	0 (0%)
diabetes*	0 (0%)	0 (0%)
general_anes	8 (0%)	0 (0%)
hypocholest*	0 (0%)	0 (0%)
postop_hematocrit	7597 (31%)	40 (17%)
postop_hemoglob	9540 (38%)	57 (24%)
preop_albumin	17468 (70%)	143 (61%)
preop_hematocrit	10220 (41%)	73 (31%)
preop_hemoglob	10709 (43%)	79 (34%)
preop_stay	1002 (4%)	14 (6%)
rheum_dx*	0 (0%)	0 (0%)
sex	13 (0%)	0 (0%)
surg_dur	614 (2%)	13 (6%)
wnd_class	1128 (5%)	34 (14%)

* Unable to determine extent of missing data; original data dictionary instructed data abstractors to send specific ICD-9 codes if condition was present or leave blank if condition was not present.

Exhibit 28. Count of missing values (with percentage)

Procedure		postop_ hematocrit	postop_ hemoglob	preop_ albumin	preop_ hematocrit	preop_ hemoglob
DH						
	HE	429 (46.4%)	429 (46.4%)	489 (52.9%)	340 (36.8%)	345 (37.3%)
	TH	47 (17.3%)	47 (17.3%)	127 (46.9%)	70 (25.8%)	70 (25.8%)
	TK	31 (13.2%)	32 (13.7%)	99 (42.3%)	41 (17.5%)	41 (17.5%)
IH						
	AP	3923 (64%)	3949 (64.5%)	2784 (45.4%)	1568 (25.6%)	1672 (27.3%)
	CA	80 (4.4%)	81 (4.5%)	418 (23.2%)	123 (6.8%)	128 (7.1%)
	HE	1690 (77.8%)	1703 (78.4%)	2173 (100%)	1424 (65.5%)	1682 (77.4%)
	TH	85 (2.9%)	514 (17.7%)	2478 (85.2%)	1249 (42.9%)	1334 (45.8%)
	TK	615 (6.9%)	1817 (20.3%)	7697 (86.2%)	4511 (50.5%)	4549 (50.9%)
VA						
	CA	78 (100%)	78 (100%)	1 (1.3%)	78 (100%)	78 (100%)
	HE	202 (85.2%)	237 (100%)	90 (38%)	22 (9.3%)	22 (9.3%)
	TH	(0%)	90 (100%)	49 (54.4%)	1 (1.1%)	1 (1.1%)
	TK	1 (0.6%)	163 (100%)	95 (58.3%)	3 (1.8%)	3 (1.8%)
VAIL						
	HE	388 (92.8%)	390 (93.3%)	418 (100%)	337 (80.6%)	336 (80.4%)
	TH	13 (8.2%)	12 (7.5%)	159 (100%)	107 (67.3%)	107 (67.3%)
	TK	55 (10.3%)	55 (10.3%)	534 (100%)	419 (78.5%)	420 (78.7%)

DH = Denver Health; IH = Intermountain Healthcare; VA = Salt Lake City VA Vail = Vail Valley Medical Center; AP = Appendectomy; CA = CABG; HE =Herniorrhaphy; TH = Total Hip Arthroplasty; TK = Total Knee Arthroplasty

In multiple imputation, missing values for any variable are predicted using existing values from other variables. The predicted values, called "imputes", are substituted for the missing values, resulting in a full dataset called an "imputed dataset." This process is performed multiple times, producing multiple imputed datasets (hence the term "multiple imputation"). Standard statistical analysis is carried out on each imputed dataset, producing multiple analysis results. These analysis results are then combined to produce one overall analysis.

Multiple imputation accounts for missing data by restoring not only the natural variability in the missing data, but also by incorporating the uncertainty caused by estimating missing data. Maintaining the original variability of the missing data is done by creating imputed values, which are based on variables correlated with the missing data and causes of data being missing. Uncertainty is accounted for by creating different versions of the missing data and observing the variability between imputed datasets. It is important to note that imputed values produced from an imputation model are not intended to be "guesses" as to what a particular missing value might be; rather, this modeling is intended to create an imputed dataset that maintains the overall variability in the population while preserving relationships with other variables. Thus, in performing multiple imputation, a researcher is interested in preserving important characteristics of the dataset as a whole (e.g., means, variances, regression parameters). Creating imputes is merely a mechanism to deliver an analysis that makes use of all possible information.

New to SAS® version 9† is the Multiple Imputation (MI) procedure,[75] which uses a random sample of missing values to account for any uncertainty from the missing data. For our project, we chose a Markov Chain Monte Carlo method to create five complete datasets, each with a slightly different value in the missing slot. We then used standard statistical analyses on the complete datasets. Also new to SAS 9 is the MIANALYZE procedure that we used to combine the five dataset-analysis results into a single inferential result.

Our next step was to randomly select 60 percent of the data to be placed in a derivation environment where we could develop the statistical models. The remaining 40 percent of the data was placed in the validation dataset, which was used for comparison and confirmation of the models built with the derivation dataset. After validation, the final statistical models used the full dataset that included both the derivation and validation datasets.

Univariate regression was used to determine the independent association of potential risk factors and SSI. The final model included risk factors with probability $\boldsymbol{P} < 0.05$, or that contributed to the predictive value of the model. We first used a binary logistic regression model to evaluate the relationship of each variable with the occurrence of an SSI. For the nonimputed variables, we used the original dataset with logistic regression. For the imputed variables, we used the five imputed datasets with a combination of logistic regression and the MIANALYSE procedure to generate results.

The risk factors for SSI were tested by including the type of procedure as a binary variable (yes/no) and the risk factors were also independently tested for each of the five different procedures (CABG, herniorrhaphies, hip arthroplasty, knee arthroplasty and appendectomies). The predictive models were created with stepwise logistic regression, using the five imputed datasets. The entry level probability for each variable was set at .2 and the probability used to keep a variable in the model was set at .25. Running logistic regression on five separate datasets resulted in five different candidate models that were mostly the same, but had a few differences. We then took any variable identified in the five multivariate models to create the final logistic regression model. We then ran the final logistic regression model on the five imputed datasets and used the MIANALYZE procedure to produce one set of results from the five iterations.

Subtask 3.4. Identify SSI Risk Factors using combined datasets from all four facilities/systems.

3.4.1. Multivariate analysis of the datasets, including each procedure as a binary variable.

During the univariate analyses of the derivation dataset, 13 different risk factors were included in the model. That analysis also included each of the five different procedures as a

---

† The data analysis for this paper was generated using SAS/STAT software, Version 9.2 of the SAS System for Windows. Copyright © 2002-2008 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA

binary variable, yes/no. Each of those 13 risk factors was then included in three different logistic regression analyses using the derivation, validation, and combined datasets (Exhibit 29, Exhibit 30, and Exhibit 31). The statistical significance of each of the 13 potential risk factors changed during each test using the three different datasets. For the derivation dataset, 6 of the 13 univariate risk factors remained significant in the model compared to only 3 in the validation set, and 7 when both the derivation and validation sets were combined. Only a history of MRSA infection and a postoperative admission within 30 days were significant in all three tests. In many cases, postoperative admission was indicative for admission due to a postoperative wound. Chronic kidney disease was significant in the derivation and combined datasets, along with an increase in the number of procedures and a low postoperative hematocrit. Male gender was only significant in the derivation analysis. CABG surgery was only found to be significant in the validation and combined datasets. A longer preoperative stay was only significant in the combined dataset.

Exhibit 29. Results of logistic regression using the 60-percent derivation set, including each of the five surgical procedures as a possible risk factor

Variable	Estimate	tValue	Probability
Intercept	-5.66922	-7.18334	0.00000
Admission transfer	0.249358	0.655373	0.51223
History or cancer	0.675356	1.234336	0.21708
Chronic kidney disease	-2.28736	-2.24028	0.02507
General anesthesia	0.419425	1.499021	0.13387
Male	0.58722	2.677562	0.00742
History of MRSA	1.47331	4.87133	0.00000
Number of procedures	0.381148	3.171798	0.00151
Postop admission	2.104624	9.115523	0.00000
Postop hematocrit	-0.04406	-2.01137	0.04496
Preop stay	0.042437	1.623837	0.10441
CABG surgery	0.24496	0.746911	0.45512
Herniorrhaphy	0.186762	0.564969	0.57209
Wound class	0.003031	0.014814	0.98818

CABG = coronary artery bypass grafting; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

Exhibit 30. Results of logistic regression using the 40-percent validation set, including each of the five surgical procedures as a possible risk factor

Variable	Estimate	tValue	Probability
Intercept	-5.9448	-7.7606	0.00000
Admission transfer	-0.46026	-0.92513	0.35490
History or cancer	-0.78305	-0.76388	0.44494
Chronic kidney disease	-0.49964	-0.91168	0.36193
General anesthesia	0.046769	0.171229	0.86404
Male	-0.08612	-0.3914	0.69550
History of MRSA	**1.149853**	**3.048832**	**0.00230**
Number of procedures	0.264643	1.861373	0.06269
Postop admission	**2.115295**	**9.348542**	**0.00000**
Postop hematocrit	-0.01706	-0.79509	0.42677
Preop stay	0.035891	1.814288	0.06963
CABG surgery	**0.725095**	**2.077166**	**0.03779**
Herniorrhaphy	0.553825	1.787663	0.07383
Wound class	0.326074	1.904355	0.05686

CABG = coronary artery bypass grafting; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

Exhibit 31. Results of logistic regression using the combined derivation and validation sets, including each of the five surgical procedures as a possible risk factor

Variable	Estimate	tValue	Probability
Intercept	-5.75478	-10.6587	0.00000
Admission transfer	-0.07527	-0.25258	0.80059
History or cancer	0.186484	0.396676	0.69161
Chronic kidney disease	-1.19328	-2.52852	0.01145
General anesthesia	0.248213	1.277382	0.20147
Male	0.254338	1.658706	0.09718
History of MRSA	**1.347607**	**5.798036**	**0.00000**
Number of procedures	**0.322013**	**3.494699**	**0.00047**
Postop admission	**2.111179**	**13.10889**	**0.00000**
Postop hematocrit	**-0.03231**	**-2.16588**	**0.03045**
Preop stay	**0.035145**	**2.315832**	**0.02057**
CABG surgery	**0.477163**	**2.015653**	**0.04384**
Herniorrhaphy	0.376815	1.673907	0.09415
Wound class	0.16983	1.300158	0.19355

CABG = coronary artery bypass grafting; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

3.4.2. Multivariate analysis of the datasets including only CABG surgery.

During the univariate analyses of the derivation dataset for CABG surgeries, seven different risk factors were included in the model. Each of those seven risk factors was then included in three different logistic regression analyses using the 60-percent derivation, 40-percent validation, and 100-percent combined datasets (Exhibits 32–34). The probability of each of the seven potential risk factors changed for each of the three different datasets. For the derivation dataset, five (transfer admission, increased BMI, history of MRSA, postoperative admission within 30 days, and longer surgery duration) of the seven univariate risk factors remained significant in the model, while only one (increased BMI) remained in the validation set. For the combined dataset, except for transfer admission, the same four factors identified using the derivation set remained significant. Of interest, postoperative admission within 30 days had a nonsignificant probability of 0.2134 in the validation set.

Exhibit 32. Results of logistic regression using the derivation dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-10.9133	-4.25511	0.00002
Transfer admission	**1.285104**	**2.234972**	**0.02542**
BMI	**0.111739**	**3.758852**	**0.00017**
History of MRSA	**2.214599**	**3.213444**	**0.00131**
Postop admission	**2.03457**	**3.397303**	**0.00068**
Preop hematocrit	0.145933	1.881071	0.06342
Preop hemoglob	-0.1963	-0.86847	0.39360
Surgery duration	**-0.00573**	**-2.02865**	**0.04249**

BMI = body mass index; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

Exhibit 33. Results of logistic regression using the validation dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-4.41194	-2.00311	0.04533
Transfer admission	-0.36612	-0.46516	0.64182
BMI	**0.087611**	**2.610583**	**0.00904**
History of MRSA	1.072824	1.320566	0.18665
Postop admission	0.646026	1.244747	0.21324
Preop hematocrit	0.027258	0.274814	0.78942
Preop hemoglobin	-0.20223	-0.70553	0.50265
Surgery duration	-0.00372	-1.3105	0.19003

BMI = body mass index; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

Exhibit 34. Results of logistic regression using the combined dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-7.01035	-4.29561	0.00002
Transfer admission	0.616473	1.420047	0.15560
BMI	**0.095185**	**4.350031**	**0.00001**
History of MRSA	**1.565599**	**3.098649**	**0.00195**
Postop admission	**1.267214**	**3.445955**	**0.00057**
Preop hematocrit	0.087034	1.579951	0.12344
Preop hemoglobin	-0.22319	-1.51617	0.14453
Surgery duration	**-0.00477**	**-2.36402**	**0.01809**

BMI = body mass index; MRSA = methicillin-resistant *Staphylococcus aureus*; Postop = postoperative; Preop = preoperative.

3.4.3. Multivariate analysis of the datasets, including only herniorrhaphy.

During the univariate analyses of the derivation dataset for CABG surgeries, seven different risk factors were included in the model. Each of those seven risk factors was then included in three different logistic regression analyses using a 60-percent derivation, 40-percent validation and 100-percent combined datasets (Exhibits 35– 37). The probability of each of the seven potential risk factors changed for each of the three different datasets. For the derivation dataset, three factors (postoperative admission within 30 days, postoperative hematocrit, and postoperative hemoglobin) remained significant in the model, and only postoperative admission remained in the validation and the combined sets. Of interest, postoperative admission within 30 days had a nonsignificant probability of 0.2134 in the validation set.

Exhibit 35. Results of logistic regression using the derivation dataset for only herniorrhaphy

Variable	Estimate	tValue	Probability
Intercept	-7.19702	-2.90005	0.00378
No. of procedures	-1.22262	-1.19116	0.23359
Postop admission	**2.105654**	**3.434398**	**0.00059**
Postop hematocrit	**-0.15235**	**-2.04515**	**0.04164**
Postop hemoglobin	**0.558391**	**2.282273**	**0.02329**
Longer preoperative stay	0.119229	1.583554	0.11330
History or rheumatism	1.80024	1.428812	0.15310
Wound class	0.812588	1.409284	0.15876

Postop = postoperative; Preop = preoperative.

Exhibit 36. Results of logistic regression using the validation dataset for only herniorrhaphy

Variable	Estimate	tValue	Probability
Intercept	-6.63084	-3.1182	0.00195
No. of procedures	0.389846	1.706724	0.08789
Postop admission	**3.430942**	**5.136659**	**0.00000**
Postop hematocrit	0.132635	1.271441	0.20613
Postop hemoglobin	-0.42173	-1.32403	0.18974
Longer preop stay	0.006993	0.13405	0.89336
History or rheumatism	0.632211	0.536855	0.59137
Wound class	0.407113	0.76837	0.44228

Postop = postoperative; Preop = preoperative.

Exhibit 37. Results of logistic regression using the combined dataset for only herniorrhaphy.

Variable	Estimate	tValue	Probability
Intercept	-6.53827	-4.0998	0.00007
No. of procedures	0.174594	0.920322	0.35743
Postop admission	**2.787336**	**6.551838**	**0.00000**
Postop hematocrit	-0.03547	-0.44784	0.66071
Postop hemoglobin	0.082922	0.331393	0.74535
Longer preop stay	0.036602	1.854843	0.06362
History or rheumatism	1.226935	1.51136	0.13070
Wound class	0.490776	1.44229	0.14923

Postop = postoperative; Preop = preoperative.

3.4.4. Multivariate analysis of the datasets including only total hip surgery.

During the univariate analyses of the derivation dataset for total hip surgeries, eight different risk factors were included in the model. Each of those eight risk factors was then included in three different logistic regression analyses using 60-percent derivation, 40-percent validation and 100-percent combined datasets (Exhibits 38–40). The probability of each of the eight potential risk factors changed for each of the three different datasets. For the derivation dataset, four factors (emergency surgery, number of procedures, postoperative admission within 30 days, surgery duration) remained significant in the model. Chronic lung disease, emergency surgery, and postoperative admission were significant in the validation set, and the probability related to surgery duration was 0.05866. For the combined dataset, the same three factors as for the validation set were significant, plus surgery duration. Because the stepwise logistic regression entry and stay probabilities were set at .20 and .25, respectively, it was possible for some variables to be nonsignificant in the derivation dataset and significant in the validation and combined datasets (chronic lung disease, for example).

Exhibit 38. Results of logistic regression using the derivation dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-3.15016	-1.31103	0.19127
Age	-0.02436	-1.36677	0.17172
Chronic lung disease	0.717234	1.299857	0.19365
Emergency surgery	**1.667783**	**2.711989**	**0.00669**
No. of procedures	**0.647188**	**2.857793**	**0.00442**
Postop admission	**1.502892**	**2.895868**	**0.00379**
Postop hemoglobin	-0.26253	-1.52152	0.12890
Preop hemoglobin	0.013404	0.231	0.82005
Surgery duration	**0.007371**	**2.714224**	**0.00665**

Postop = postoperative; preop = preoperative

Exhibit 39. Results of logistic regression using the validation dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-5.30978	-2.3102	0.02089
Age	-0.01718	-1.01601	0.30962
Chronic lung disease	**1.181684**	**2.524745**	**0.01158**
Emergency surgery	**2.01501**	**3.420056**	**0.00063**
No. of procedures	-0.58576	-0.61154	0.54084
Postop admission	**1.962564**	**3.988843**	**0.00007**
Postop hemoglobin	0.079368	0.560087	0.57554
Preop hemoglobin	0.019021	0.361141	0.72071
Surgery duration	0.002467	1.890769	0.05866

Postop = postoperative; Preop = preoperative

Exhibit 40. Results of logistic regression using the combined dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-4.28495	-2.91129	0.00361
Age	-0.02085	-1.74905	0.08028
Chronic lung disease	**0.930147**	**2.645746**	**0.00815**
Emergency surgery	**1.730828**	**4.182002**	**0.00003**
No. of procedures	0.410227	1.931115	0.05347
Postop admission	**1.728514**	**4.971104**	**0.00000**
Postop hemoglobin	-0.09155	-0.89798	0.36925
Preop hemoglobin	0.014634	0.464605	0.64256
Surgery duration	**0.003186**	**2.997669**	**0.00272**

Postop = postoperative; Preop = preoperative

3.4.5. Multivariate analysis of the datasets including only total knee surgery.

During the univariate analyses of the derivation dataset for total knee surgeries, only five different risk factors were included in the model. Each of those five risk factors was then included in three different logistic regression analyses using a 60 percent derivation, 40 percent validation and 100 percent combined datasets (Exhibit 41, Exhibit 42, and Exhibit 43). The probability of each of the five potential risk factors changed for each of the three different datasets. For the derivation dataset, three univariate risk factors (history of MRSA, number of procedures, and postoperative admission within 30 days) remained significant in the model. Three of the five (history of MRSA, postoperative admission within 30 days, and preop hemtocrit) were significant in the validation set. When the five univariate risk factors were tested with the combined dataset, all five (including male gender) were significant.

Exhibit 41. Results of logistic regression using the derivation dataset for only total knee surgery

Variable	Estimate	tValue	Probability
Intercept	-5.75277	-4.00541	0.00007
Male	0.680791	2.12279	0.03377
History of MRSA	**1.752457**	**4.314211**	**0.00002**
No. of procedures	**0.748074**	**4.581106**	**0.00000**
Postop admission	**2.233521**	**5.919509**	**0.00000**
Preop hematocrit	-0.03908	-1.20797	0.22750

MRSA = methicillin-resistant *Staphlococcus aureus*; Postop = postoperative; Preop = preoperative

Exhibit 42. Results of logistic regression using the validation dataset for only total knee surgery

Variable	Estimate	tValue	Probability
Intercept	-3.80701	-2.42999	0.01528
Male	0.558903	1.518404	0.12891
History of MRSA	**1.306538**	**2.061606**	**0.03925**
No. of procedures	0.383498	1.838999	0.06592
Postop admission	**2.166304**	**5.228965**	**0.00000**
Preop hematocrit	**-0.07291**	**-1.99583**	**0.04627**

MRSA = methicillin-resistant *Staphlococcus* aureus; Postop = postoperative; Preop = preoperative

Exhibit 43. Results of logistic regression using the combined dataset for only total knee surgery

Variable	Estimate	tValue	Probability
Intercept	-4.99176	-4.53387	0.00001
Male	**0.623433**	**2.577626**	**0.00995**
History of MRSA	**1.593127**	**4.739078**	**0.00000**
No. of procedures	**0.584536**	**4.752973**	**0.00000**
Postop admission	**2.179504**	**7.873071**	**0.00000**
Preop hematocrit	**-0.05078**	**-2.00126**	**0.04697**

MRSA = methicillin-resistant *Staphlococcus* aureus; Postop = postoperative; Preop = preoperative

3.4.6. Multivariate analysis of the dataset including only appendectomy surgery at Intermountain Healthcare.

During the univariate analyses of the derivation dataset for appendectomy surgeries, seven different risk factors were included in the model. Each of those seven risk factors was then included in three different logistic regression analyses using the derivation, validation, and combined datasets (Exhibits 44–46). The probability of each of the seven potential risk factors changed during each test, using the three different datasets. For the derivation dataset, only two (postoperative admission within 30 days and postoperative hematocrit) of the seven univariate risk factors remained significant in the model, and only one (postoperative admission) remained in the validation set. For the combined datasets, the same two as the derivation set remained significant. Again, only postoperative admission within 30 days was significant in all three tests.

Exhibit 44. Results of logistic regression using the derivation dataset for only appendectomy surgeries.

Variable	Estimate	tValue	Probability
Intercept	-3.45066	-1.04967	0.31289
History of cancer	0.896303	0.974859	0.32970
Emergency surgery	-0.62318	-0.96643	0.33384
Male	0.488666	0.768574	0.44219
Postop admission	**2.453247**	**3.652926**	**0.00026**
Postop hematocrit	**-0.17629**	**-2.12084**	**0.04042**
Postop hemoglobin	0.272209	0.822076	0.42317
Preop hemoglobin	-0.0374	-0.26214	0.79751

Postop = postoperative; Preop = preoperative

Exhibit 45. Results of logistic regression using the validation dataset for only appendectomy surgeries

Variable	Estimate	tValue	Probability
Intercept	-4.59592	-2.06151	0.04255
History of cancer	0.855832	0.736814	0.46124
Emergency surgery	0.655977	1.172465	0.24101
Male	0.586734	1.019488	0.30824
Postop admission	**2.572595**	**4.931689**	**0.00000**
Postop hematocrit	-0.15581	-1.68912	0.09748
Postop hemoglobin	0.310595	1.163695	0.24698
Preop hemoglobin	-0.04081	-0.31694	0.75426

Postop = postoperative; Preop = preoperative

Exhibit 46. Results of logistic regression using the combined dataset for only appendectomy surgeries

Variable	Estimate	tValue	Probability
Intercept	-3.90766	-2.4378	0.01642
History of cancer	0.749083	1.068061	0.28549
Emergency surgery	0.130644	0.331969	0.73991
Male	0.619657	1.471467	0.14148
Postop admission	**2.515271**	**6.19406**	**0.00000**
Postop hematocrit	**-0.15793**	**-2.74004**	**0.00744**
Postop hemoglobin	0.304609	1.439065	0.16263
Preop hemoglobin	-0.07631	-0.6829	0.51247

Postop = postoperative; Preop = preoperative

Subtask 3.6. Summary

3.6.1. Identified SSI risk factors using datasets from all four facilities/systems.

This study examined a large number of potential risk factors contained in electronic medical records from four different facilities/systems in an effort to identify predictors for SSI. The potential risk factors were identified through surgeons' experience and an extensive literature review for any potential risk factors for any surgical procedure during the past 10 years. Four

different surgical procedures (herniorrhaphy, CABG, total hip, total knee) were included in this study from all four facilities/systems, and appendectomy was also included from Intermountain Healthcare. Analyses of six different data partitions were conducted which included all data and each of the different surgical procedures as a potential risk factor, and then data from each surgical procedure separately. For each of the different data partitions, three iterations were conducted which included a 60-percent random derivation set, the remaining 40-percent validation set, and then the combined dataset.

During the analyses of the three iterations that included the different surgical procedures as binary risk factors and the three iterations for the individual tests of each of the five different surgical procedures, 21 (64 percent) of the 33 potential risk factors tested were identified during the univariate analyses during at least one iteration. However, only 13 of the 21 were found to be significant in at least one of the derivation, validation, or combined multivariate logistic regression models. While herniorrhaphy was included as a risk factor during the univariate analysis that included surgery type as a binary variable, only CABG surgery was statistically significant in the validation and combined regression models. The most common risk factor (identified during 16 of the 18 different iterations) was postoperative admission within 30 days. Rather than a preoperative risk factor, this finding was indicative of the need for hospitalization for postoperative wound treatment. One might ask why was postop admission within 30 days included in the analysis for this study. First, it is a common risk factor or "trigger" used by many facilities that rely on manual SSI surveillance, including postdischarge phone calls to patients, for postdischarge SSI identification. This study shows that it is still a reliable method to identify SSIs. Second, since it was the most common risk factor identified in this study, that led us to feel the statistical analysis was working appropriately. If postoperative admission was not identified as a risk factor, it would have immediately led us to question our results and methods. The next most common risk factor was history of MRSA (identified seven times) followed by postop hematocrit (six times), number of procedures (five times), surgery duration (four times), and increased BMI and postop hemoglobin at three times each. It was interesting to note that the significant risk factors almost always varied between the derivation, validation, and combined datasets.

This study indicates that risk factors for surgical site infections vary by the type of procedure. The risk models generated for each of the five different types of surgical procedure and the inclusion of procedure type varied. The number of risk factors included in the different models also varied from 5 to 13 in this study. We also found that the risk factors for total hip surgery differed from those for total knee surgery. Thus, total hip surgeries need to be compared with total hip surgeries, and not just the group of other orthopedic surgeries.

3.6.2. Identified SSI risk factors, using data only from Intermountain Healthcare, and including eight additional potential risk factors

The Intermountain data contributed 87 percent of the total dataset—including 57 percent of the data for herniorrhaphy, 85 percent of the total hip replacement data, 90 percent or the total knee replacement data, 95 percent of the CABG data, and all of the appendectomy data. The analysis with the Intermountain data only demonstrated the impact the other data had when included in the total dataset. It also reinforced the conclusion that risk factors can change, depending on the facilities data. As before, postop admission was the most common factor identified and was significant in all 18 iterations, followed again by history of MRSA

(significant nine times). This was possibly influenced by Intermountain's ability to monitor postop admissions across all 22 hospitals and its enterprise-wide MRSA surveillance network that monitors all MRSA patient movement throughout the system. When each of the five surgical types was included in the analyses, only herniorrhaphy was significantly associated with SSI. While herniorrhaphy was identified in the univariate analysis of the total dataset, it was not significant in either the derivation or validation analyses, whereas CABG was significant in both the validation and combined analyses. Thus, the 13 percent of the data from the other facilities did have an impact on the risk factors identified by the univariate analyses and the significance of the risk factors in the multiple regression analyses.

The inclusion of the eight new, potential risk factors in the analysis of only Intermountain data did not result in any major differences in the identification of risk factors. Charlson score, number of surgeons, and preop glucose were the only new risk factors that were identified in any of the univariate analyses, and only Charlson score and number of surgeons were ever found to be significant in the multiple regression. The number of procedures during the same surgery (concurrent procedures) was identified as a significant risk factor from the total dataset. It was significant during the derivation and combined analyses, which included the five surgical procedures as risk factors. The number of procedures was not identified in the univariate analysis using the Intermountain data only, and seems to have been replaced by the number of surgeons. The number of different surgeons would seem to be associated with the number of different procedures performed, and probably kept the number of procedures from remaining in the model.

3.6.3. Conclusion.

There is not a single set of risk factors that can be used to predict SSI across all types of surgical procedures or facilities. This study found that SSI risk factors are dependent on the type of surgical procedure. Thus, SSI rate comparison needs to be at the surgical procedure level and not the surgical service level (i.e., orthopedics, general surgery, thoracic, etc.). In addition, SSI rates should also be compared at the facility level against its own baseline rates. When compared at that level, the sets of risk factors identified for the five different surgical procedures in this study could be used to identify changes in SSI rates over time, and differences between surgeons.

Chapter 4. Assessing Surgeon Acceptance of Risk Adjustment Models

The originally proposed approach for Chapter 4 was to solicit input from surgeons and IP nurses on the types of risk factors that should be examined. Due to delays in Institutional Review Board (IRB) approval at Intermountain for the nursing focus group, we used this opportunity (with permission from AHRQ and CDC) to repurpose the nursing focus group. Surgeon focus group inputs, as described below, were used to inform Chapter 3. Results of the nursing focus groups are being used to support adoption of the developed tools by offering greater insight into the decision to adopt and issues/challenges around implementation of the Chapter 2 surveillance tool.

Subtask 4.1. Identify a Representative Group of Active Surgeons in the United States.

4.1.1. Identify surgeons to participate in a focus group and in-depth discussion

Surgeons were identified based on their project-relevant professional expertise, and were selected from multiple health care settings and systems in order to maximize the representative nature of the focus group participants. Walter L. Biffl, M.D., recruited participant surgeons through use of a national conference-based professional networking strategy. National meeting organizers for the 5th annual Academic Surgical Congress were contacted to solicit their facilitation and ensure support, as well as to secure space to conduct the focus group adjacent to the national meeting on February 3, 2010, in San Antonio, TX.

4.1.2. Identify nurses to participate in focus groups and in-depth discussion

Two focus groups were conducted. The first was scheduled for Denver, CO, and infection control and/or infection preventionist nurses regularly working in SSI surveillance were recruited from the Mile High chapter of the Association for Professionals in Infection Control and Epidemiology (APIC). The second focus group was scheduled for Salt Lake City, UT, and infection control and/or infection preventionist nurses regularly working in SSI surveillance were recruited from the APIC Infection Control Association. Participants were given a $35 honorarium for participation.

Subtask 4.2. Develop a Mechanism to Ascertain Opinions and Ideas From Surgeons and IP Nurses About Current Risk-Adjustment Models and Proposed Changes

4.2.1. Develop focus group guides

Three distinct focus group guides were developed, one for use with each of the above-mentioned, homogeneous groups.

Surgeon focus group. The guide was designed to include a series of tailored, open-ended questions regarding surgeons' attitudes towards current risk adjustment models; the variables that they consider important among their patients, and the appropriateness and relative risks of those variables; and ideas on optimizing feedback to enhance performance. The guide was developed through a process of iterative review by the project team, and was reviewed by the Colorado Multiple Institutional Review Board (COMIRB) prior to use (see Appendix N).

Nursing focus groups. We developed two nursing focus group guides with the intent to explore the decision to adopt the e-detection tool and issues around its implementation (see Appendix O). A brief educational symposium on SSI detection and overview of our tool (developed in Chapter 2) was followed by an open discussion with infection control/prevention nurses—one group in Denver and one in Salt Lake City. Results were used to support development of an implementation manual for dissemination.

4.2.2. Secure IRB approval

IRB approval for the surgeon focus group was secured by Denver Health on January 29, 2010, from COMIRB. An IRB exemption was secured by Intermountain Healthcare on January 17, 2011, for the nursing focus groups.

4.2.3. Secure Office of Management and Budget clearance or a clearance exemption

Consultation was held with AHRQ's Task Order Officer and the Office of Management and Budget (OMB) Liaison regarding the type of clearance exemption that would be required for the project. It was determined that OMB review and clearance was not required. No clinical exemption was necessary, because the clinical data used for the project was already required to be collected for other purposes. The new data collected through focus group sessions was not subject to OMB clearance requirements, as each focus group was designed to explore a topic area distinct from the other groups, and no group had more than nine participants. No further review was deemed necessary.

4.2.4. Conduct focus groups

Focus groups were conducted using a team approach, involving both a moderator and a trained qualitative researcher. The moderator promoted interaction and guided the discussion to ensure that the focus remained on the topic of interest, order was maintained, and all participants were engaged. The participation of an observing researcher in addition to the moderator allowed for an accurate record to be made without interrupting the flow of discussion.

Each focus group was documented through summary notes taken by the researcher and through audio recording of the session. Documentation included both a record of spoken responses and observation of group members' interactions. The use of redundant documentation methods to augment moderator and researcher recollection ensures the most complete set of data for analysis. Participant initials were recorded with individual responses to ensure accuracy of data analysis (i.e., not attributing a response to multiple individuals when a single individual raised the same point several times).

Surgeon focus group. Six surgeons in addition to the facilitator participated in the focus group conducted on February 3, 2010, adjunct to the 5th annual Academic Surgical Congress in

San Antonio, TX. Participants were recruited by the facilitator based on their presence and involvement at the national meeting and interest in the topic under discussion. Surgeons represented multiple health system types and surgical specialties, as described in Exhibit 47 below:

Exhibit 47. Surgeon specialties and health system affiliations

Surgeon ID	Surgical Specialty	Health System Type
Surgeon #01	General Surgery	Academic / Private
Surgeon #02	General Surgery	Academic
Surgeon #03	General Surgery	VA
Surgeon #04	General Surgery – Trauma/Critical Care focus	Academic / Safety Net
Surgeon #05	General Surgery – Trauma/Critical Care focus	Private
Surgeon #06	Surgical Oncology	Academic / Safety Net

Nursing focus groups. Five infection control nurses in addition to the facilitator and an observing notetaker participated in the first focus group conducted on February 2, 2011, in Denver. Participants were recruited by the facilitator based on their involvement with the Mile High chapter of APIC and their interest in the topic to be discussed.

Eight infection control nurses in addition to the facilitator participated in the second focus group conducted on February 25, 2011, in Salt Lake City. Participants were recruited from various hospitals from within Intermountain Healthcare.

4.2.5. Conduct followup discussions with focus group participants

No followup discussions were deemed necessary.

Subtask 4.3. Compile Results of Focus Groups to Make Recommendations of Surgeon, Nurse Perspectives on Risk-adjustment Models for SSI

4.3.1. Independent analysis of focus group and discussion data

Surgeon focus group. Focus group data were analyzed through an inductive approach that used an open, heuristic coding process to identify initial topics mentioned by participants. Individual topics were further categorized, based on the number of participants who conveyed agreement with the concept being discussed. A topic was identified as a theme based on the mention of or agreement with an item by three or more individual participants.

Data were reviewed to the saturation point and discussed with the focus group facilitator and subject expert to ensure the most comprehensive identification of patterns. Topics identified as duplicative were combined into a single occurrence, and themes identified from the comprehensive topic list based on the number of individual mentions.

Nursing focus groups. Data from the first focus group were inductively analyzed by the group facilitator and content expert to elicit an understanding of current infection surveillance processes and to assess potential alteration of the standard process flow by the presence of an electronic SSI surveillance tool. A summary of key points emerging from the discussion was used to inform the second nursing focus group, centered on implementation, and conducted in Salt Lake City on February 25, 2011. Results from the second focus group informed the development of the implementation manual (Appendix R).

4.3.2. Research team meeting and consensus discussion

Results of the surgeon focus group were discussed on team calls and at the in-person meeting in October, 2010. Appendix P provides risk factor inputs from the surgeon focus group.

4.3.3. Summary paper preparation and presentation

Surgeon focus group. Draft and final versions of the results of the content analysis of focus group data were presented in report summary form to the research team, the TOO, and the Technical Experts. A copy of the final version is attached as Appendix Q. Results were used to inform the selection of common factors for Chapter 3, and to gain insight into how surgeons might use the risk adjustment tool and, thus, how dissemination and adoption might better be promoted. A limitation of these results is that the majority of focus group surgeon participants were primarily expert in general surgery. We acknowledge the possibility that additional risk factors specific to CABG or hip and knee arthroplasty procedures might have been identified in a more diverse group.

The consensus among surgeon participants was that current models for SSI risk assessment were inadequate for their needs. Risk models either inappropriately accounted for the factors they included or included an excessive number of factors, such that items of actual significance were obscured. The surgeons expressed desire for the development of new models based on specific patient factors that are identified as significant in affecting risk rates.

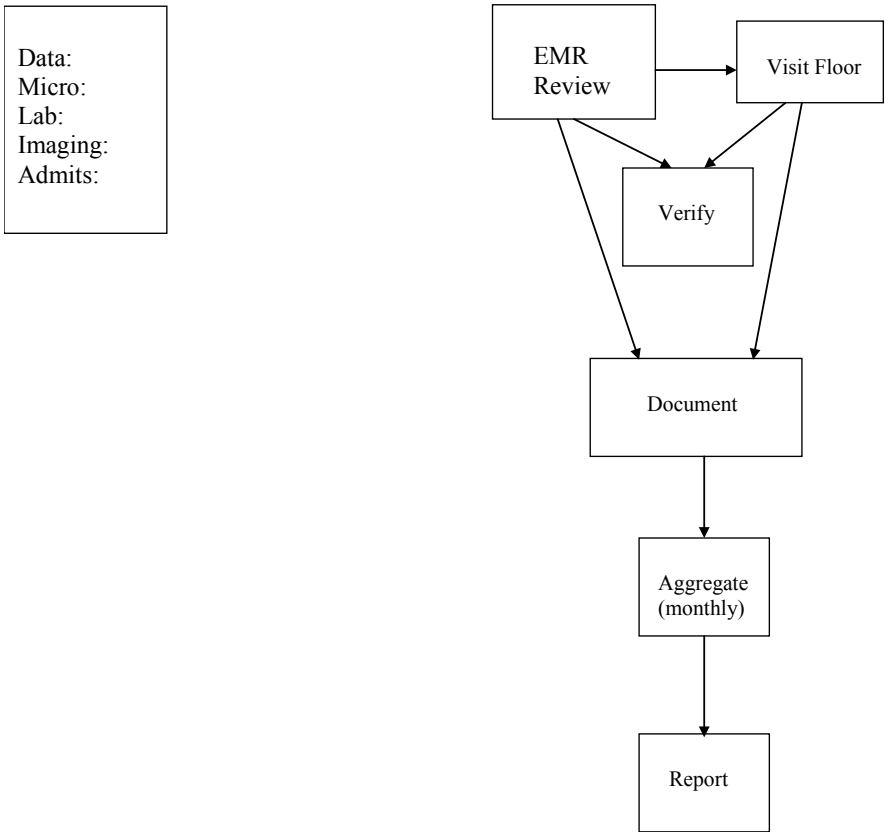
Surgeons agreed that infection rate assessments varied, based on items such as whether rates were determined based on process or outcome factors, what methods of documentation were used to report rates between private and public settings, whether or not rate data provided to an membership-based analysis database by participant hospitals are reflective of all populations, and whether or not there was variance in the interpretation of how risk factor measures were defined.

Likewise, risks were determined to vary with some factors to a degree such that different broad categories of risk might be considered, such as risks for emergency surgery patients versus those for elective surgery patients; risks for a patient with managed comorbidities versus those for patients with poorly managed or undocumented comorbidities; risks for patients in compliance with medical recommendations versus risks for noncompliant patients; and whether—in some cases—scheduling considerations, operation timing, or the risk to a patient that might result from a delayed operation outweighed the risk of infection resulting from the operation itself.

Finally, surgeons suggested approaches for improving risk assessment and management, such as giving provider-level feedback in a timely fashion; increasing risk awareness by drawing attention to measurement and tracking of risk factors; and intervening on one or more risk factor

variables, based on a patient's risk level, instead of taking a "one size fits all" approach to risk.

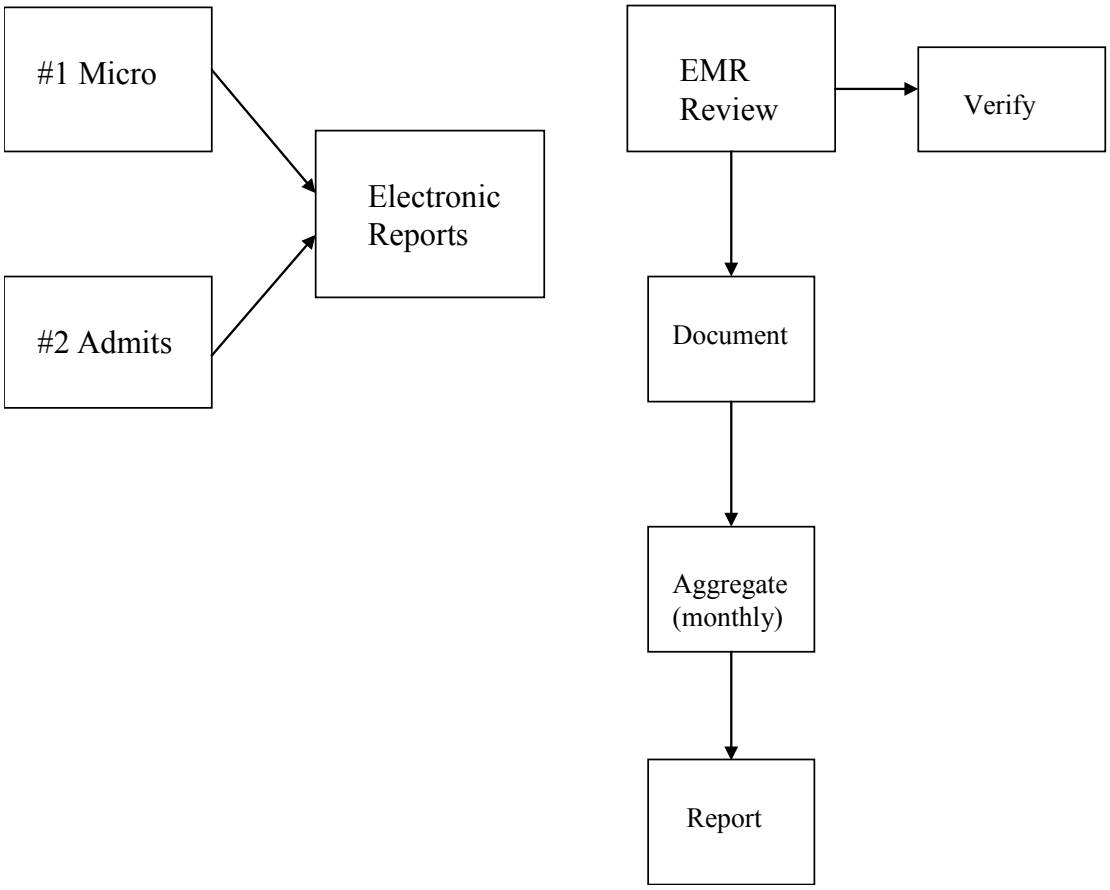
Exhibit 48. Original manual process flow for surgical site infection surveillance



EMR = electronic medical record

Nursing focus group results. The first of two focus groups was held on February 2, 2011. The purpose of this focus group was to cultivate an in-depth understanding of the decision to adopt an electronic SSI surveillance tool. A strawman flow diagram of unassisted SSI surveillance activities for IPs was amended as part of this process (see Exhibit 48 for the generic flow chart and Exhibit 49 for the amended flow chart). Overall, there was general acceptance and willingness to use an electronic cognitive-support tool. Themes gleaned from this focus group informed the second focus group on implementation, held in Salt Lake City on February 25, 2011.

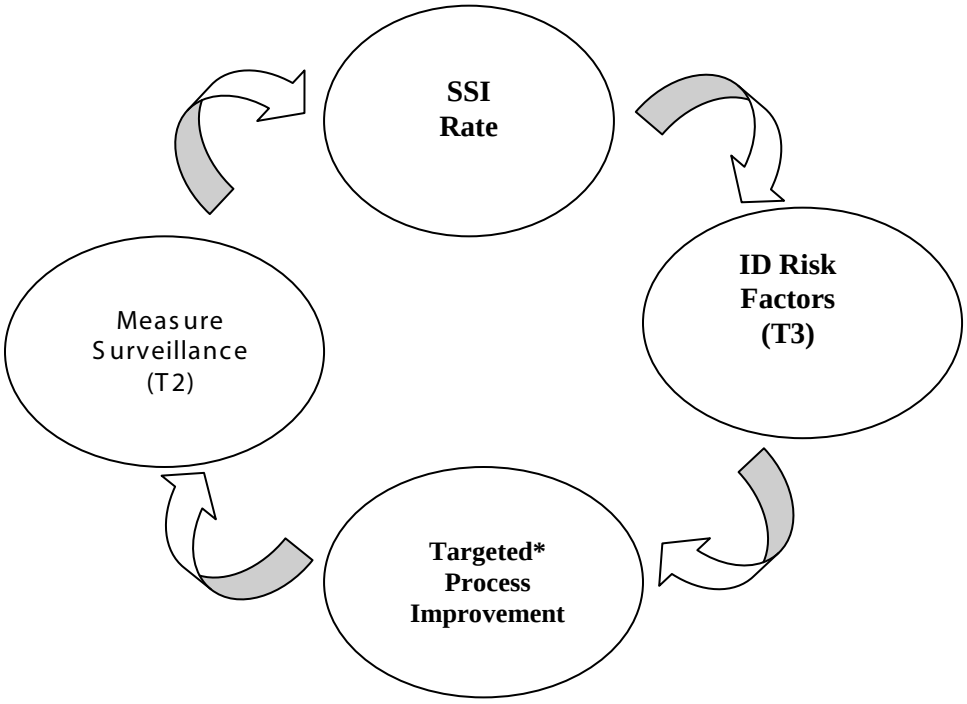
Exhibit 49. Revised manual process flow for surgical site infection surveillance



EMR = electronic medical record

The generalized model for continuous quality improvement in reducing SSIs is provided in Exhibit 50. Results from the focus groups, together with the outputs from Chapters 2 and 3, were used to develop an implementation manual (see Appendix R). The manual incorporates tools and strategies developed in both Tasks 2 and 3. As such, the manual is intended to describe how Chapter 2's measures of surveillance allow you to determine if what you did as a result of Chapter 3's risk factor assessment is working or not, as depicted in Exhibit 50.

Exhibit 50. Model for continuous quality improvement in reducing surgical site infections



***Patient, surgery, environmental
ID = identify; SSI = surgical site infection

Subtask 5.6. List of all Resulting or Anticipated Scientific Presentations and Publications From the Project

The broad-based dissemination of research results and tools developed through ACTION initiatives is an objective of primary importance. Throughout the duration of this project, our team has worked to identify a wide range of potential audiences and opportunities for dissemination at both local and national levels.

In addition to the user manual described in Chapter 4, two poster presentations and one oral presentation have been given at national conferences, and a draft manuscript has been developed for submission to a peer-reviewed journal. An additional manuscript is currently in the initial design phase. Each of these activities is further described below.

Poster presentations.

- *Surgeons' Acceptance of Surgical Site Infection Risk Adjustment Models.* This poster was presented at the Society for Healthcare Epidemiology of America (SHEA), April 1 – 4, 2011, in Dallas, TX. SHEA's mission is the prevention and control of infections in health care settings. SHEA is dedicated to advancing the science and practice of

health care epidemiology and to the prevention and control of morbidity, mortality, and costs connected to health care-associated infections. Copies of the abstract and poster are included as Appendix S.

- *Performance of Two Surgical Site Infection Risk Stratification Models for Predicting Infection Risk in Publicly Reported Data from a Safety Net Hospital*. This poster was presented at the Surgical Infection Society (SIS) annual conference, May 11-14, 2011, in Palm Beach, FL. The mission of the SIS is "to educate health care providers and the public about infection in surgical patients and promote research in the understanding, prevention and management of surgical infections." Society members include physicians, nurses, allied health personnel, scientists and others with an interest in surgical infections. Copies of the abstract and poster are included as Appendix T.

Oral presentations. *Improving the Measurement of Surgical Site Infection (SSI) Risk Stratification and Outcome Detection.* This oral presentation was given at the 2nd Annual HAI Investigators' Meeting on September 18, 2011, which was held adjunct to the AHRQ 2011 Annual Conference in Bethesda, MD. A copy of this presentation is included as Appendix U.

Manuscripts.

- "Screening for Surgical Site Infections by Applying Classification Trees to Electronic Data." This draft manuscript, focused on the results from and methods utilized in Chapter 2 of this project, is being prepared for submission to the *Journal of the American Medical Informatics Association* (*JAMIA*). *JAMIA* is the premier peer-reviewed journal of the American Medical Informatics Association (AMIA), and as such focuses on biomedical and health informatics topics that encompass the full breadth of the field, including clinical care and research, translational science, implementation science, imaging, education, consumer health, public health, and policy. Over 4,000 health care professionals are affiliated with AMIA. A copy of the draft manuscript is included as Appendix V.
- A second manuscript focusing specifically on the results from and methods utilized for Chapter 3 of this project is currently in development. We anticipate finalizing this paper for submission to an open-access, peer-reviewed journal before the end of the calendar year, extending the discussion of methodological challenges encountered and solutions for overcoming these.

Implementation manual. An implementation manual was developed to provide specific instructions on running the algorithm and conducting site-specific testing of that application. An illustrated case study for adapted use of the tool is included. We learned in Chapter 4 from focus groups with IP nurses that the Chapter 2 algorithm could serve as a valuable tool to allow them to work more efficiently, by reducing the number of unnecessary chart reviews and allowing them to more effectively concentrate on prevention activities. We then culled results from Chapter 2, together with an out-of-scope case study (see the following section, Project Expansion), to enrich the opportunity to disseminate the results of our work. The implementation manual is provided as Appendix R. The target audience for the manual is IP staff working in hospital settings. The DH-led team will work with the AHRQ Office of Communications and Knowledge Transfer (OCKT) to identify relevant dissemination channels for the manual.

Project expansion. The DH-led research team conducted two relevant analyses that extend our Chapter 2 and Chapter 3 work. These are summarized in Chapter 5.

Chapter 5. Project Expansion

Expanding Task 2. Application of a Modified Algorithm at Denver Health

In addition to the activities described in previous chapters, the infection control team at Denver Health (DH) sought to further adapt, tailor, and validate the electronic detection algorithm created in Chapter 2 for use in everyday surveillance of surgical site infections at DH, to reduce the burden of chart review while also identifying a high percentage of SSI. The mandate for the Expansion algorithm was to maximize sensitivity at the expense of specificity, while realizing a meaningful reduction in the chart review burden experienced by infection control staff. "Meaningful" was loosely estimated as at least a 50 percent reduction while maintaining at least 95 percent specificity. The team focused specifically on the loose algorithm rule, as it was considered the more sensitive model in the main project.

DH's Infection Prevention Data Manager, Bryan Knepper, generated a retrospective cohort of procedures, including associated SSI as defined by NHSN definitions, using DH surveillance data from 2007-2010. Procedures included hip and knee arthroplasty, abdominal and vaginal hysterectomy, spinal fusion, craniotomy, and herniorrhaphy.

The suggested algorithm components generated in Chapter 2 were reevaluated in a DH-specific context. Mr. Knepper, who was not an original member of the project team, met with a group of clinical professionals to discuss the loose rule parameters that would be most useful—given the DH patient population, clinical practices and prescribing and ordering tendencies—to determine which variables were most easily and reliably obtained through the DH centralized data warehouse. As a result of these discussions, along with subsequent validation of each parameter using an expanded dataset from DH, some factors from the main project's loose algorithm were incorporated faithfully into the Expansion model while other parameters were deleted, altered or expanded. "NE_N", "ESR", and "postopabx" were deleted. "Postopcx" was incorporated faithfully. The cut-point for "wbc" was changed from 9,000 cells/ml to 10,000 cells/ml. The biggest change to the models generated through the main project was to the parameter "postopadmit". The Expansion algorithm was trained to look for both outpatient visits and inpatient admissions. Large gains were realized in sensitivity, with substantial decreases to specificity. To offset decreased specificity, only visits/admissions associated with a specific list of ICD-9 codes[76] were searched for. The listed ICD-9 codes are generally associated with infection, with some directly related to surgical site infection (see Exhibit 51).

Variables included in the DH modified algorithm were leukocytosis (white blood cell count > 10,000 cells/mL), a culture (regardless of result), or a followup visit associated with any of a list of SSI-related ICD-9 codes (see Exhibit 51).

Exhibit 51. ICD-9 codes associated with followup visits in modified Denver Health algorithm

Code	Description
338.18	Other acute postoperative pain
998 codes	Postoperative complications
711 codes	Arthropathy associated with infection
996 codes	Postoperative complications
680 codes	Carbuncle and furuncle
682 codes	Cellulitis and abscess
789 codes	Other symptoms involving abdomen and pelvis
V51.8	Aftercare involving the use of plastic surgery
V71.89	Observation for other specified suspected conditions

From this, 2,179 procedures were included in the cohort. Sixty procedures were associated with SSI after manual chart review using NHSN methodology (Exhibit 52).

Exhibit 52. Modified Denver Health algorithm performance data

Procedure Type	Total Procedures	SSI	Procedures Designated for Review by Algorithm	Percent of Total Procedures
CRAN	306	14	190	62%
FUSN	191	8	88	46%
HERN	760	10	112	15%
HPRO	303	8	134	44%
HYST	159	12	67	42%
KPRO	349	7	182	52%
VHYS	111	1	31	28%
Overall	2,179	60	804	37%
Estimated Minutes Per Procedure			20	
Number of Procedures Removed from Review			1,375	
Time Saved (8 hour days)			57.3	
Sensitivity			100%	
Specificity			72%	

Results. The modified algorithm flagged 804 procedures (37 percent of total charts) for review. The percent of total procedures flagged for review varied by procedure type, and ranged from 15 percent (herniorrhaphy) to 62 percent (craniotomy). The modified algorithm achieved 100 percent sensitivity and 72 percent specificity in detecting SSI validated on 4 years of our manual SSI surveillance data using NHSN methodology.

Potential for savings. Over the four year period, 1,375 unnecessary chart reviews would have been avoided without sacrificing detection of a single SSI. Assuming 20 minutes per chart for manual review, 57 full (8-hr.) days of chart review would have been eliminated using the algorithm for surveillance of SSI in hip and knee arthroplasty, abdominal and vaginal hysterectomy, spinal fusion, craniotomy, and herniorrhaphy.

Conclusions. DH was able to successfully adapt, tailor, and validate the electronic detection algorithm (generated in Chapter 2) to determine SSI rates for an expanded set of surgical procedures, including hip and knee arthroplasty, abdominal and vaginal hysterectomy, spinal fusion, craniotomy, and herniorrhaphy at Denver Health. The modified algorithm was tailored to our setting, to be 100 percent sensitive while still reducing overall chart review burden by 63 percent.Over a 4-year period, this would have saved 57 full days of chart review at our institution, allowing more time for education and other active infection prevention interventions. The successful adaptation of the Chapter 2 algorithm to maximize sensitivity for utilization with both inpatient and outpatient visits instead of solely for postoperative admissions by an individual who was uninvolved in its development demonstrates the potential for translation into practice on a broad scale, which was one of the goals of this project.

Expanding Task 3. Testing additional risk factors using uniquely available Intermountain Healthcare data

As an expansion of Subtask 3.5, we identified SSI risk factors using only the dataset from Intermountain that included eight additional potential risk factors. Since 87 percent of the data came from Intermountain, we also analyzed the Intermountain data alone to see if we could detect any differences in risk factors for SSI. The same statistical methods were used as for the previous analysis that included all the data from the three other facilities plus the Intermountain data. However, this analysis also included eight additional potential risk factors that were electronically available in the Intermountain EDW (Exhibit 24), but not available at all of the other three facilities. Thus, 41 potential risk factors were included.

E3.5.1. Multivariate analysis of the datasets, including each procedure as a binary variable.

During the univariate analyses of the derivation dataset, 11 different risk factors were included in the model compared to 13 for the total dataset. As before, that analysis also included each of the five different procedures as a binary variable (yes/no). Each of those 11 risk factors was then included in three different logistic regression analyses using a 60 percent derivation set, a 40 percent validation set and then the combined datasets (Exhibit 53, Exhibit 54, and Exhibit 55). Again, the significance of each of the 11 potential risk factors changed during each test using the three different datasets. For the derivation dataset, eight of the 11 univariate risk factors remained significant in the model compared to five in the validation set and then 9 when both derivation and validation sets were combined. Not only were history of MRSA infection and a postoperative admission within 30 days significant in all three tests as in the previous analysis, but chronic kidney disease, preop hemoglobin, and herniorrhapy were also significant. Two of the eight new risk factors made it into the multivariate analysis—number of surgeons and preop glucose. While preop glucose was not found significant in any of the three analyses, number of surgeons was significant in the derivation and combined analyses. Compared to the previous combined analyses with the total dataset, the combined analysis with only the Intermountain data identified additional risk factors including emergency surgery, being male, number of surgeons, preop hemoglobin, surgery duration, and herniorrhaphy, while number of procedures, postop hematocrit, and CABG surgery were no longer significant. As before, postoperative admission was indicative for admission due to a postoperative wound.

Exhibit 53. Results of logistic regression using the 60% Derivation Set, including each of the 5 surgical procedures as a possible risk factor.

Variable	Estimate	tValue	Probability
Intercept	-5.47	-4.82	0.0000
Emergency surgery	0.74	3.05	0.0023
Chronic kidney disease	-1.84	-2.49	0.0130
Male	0.52	2.37	0.0180
History of MRSA	1.50	4.74	0.0000
No. of surgeons*	0.79	2.39	0.0168
Postop admission	2.58	9.73	0.0000
Preop glucose*	0.00	0.55	0.5893
Preop hemoglobin	-0.24	-4.04	0.0001
Preop stay	0.02	0.82	0.4096
Surgery duration	0.00	1.61	0.1073
Herniorrhaphy	0.88	2.90	0.0037

*Additional Intermountain risk factor.

Exhibit 54. Results of logistic regression using the 40% validation set, and including each of the five surgical procedures as a possible risk factor.

Variable	Estimate	tValue	Probability
Intercept	-4.97	-3.66	0.0003
Emergency surgery	0.01	0.04	0.9686
Chronic kidney disease	-1.54	-2.05	0.0400
Male	0.13	0.60	0.5510
History of MRSA	1.12	3.10	0.0020
No. of surgeons*	0.83	1.82	0.0695
Postop admission	1.95	8.07	0.0000
Preop glucose*	0.00	-1.00	0.3223
Preop hemoglobin	-0.17	-2.62	0.0090
Preop stay	0.04	1.13	0.2574
Surgery duration	0.00	3.07	0.0021
Herniorrhaphy	0.84	2.76	0.0059

*Additional Intermountain risk factor.

Exhibit 55. Results of logistic regression using the combined derivation and validation sets and including each of the five surgical procedures as a possible risk factor.

Variable	Estimate	tValue	Probability
Intercept	-5.41	-6.37	0.0000
Elective surgery	0.42	2.25	0.0246
Chronic kidney disease	-1.68	-3.20	0.0014
Male	0.31	1.99	0.0469
History of MRSA	1.35	5.70	0.0000
No. of surgeons*	0.80	3.05	0.0023
Postop admission	2.26	12.74	0.0000
Preop glucose*	0.00	-0.27	0.7842
Preop hemoglobin	-0.19	-4.23	0.0000
Preop stay	0.04	1.51	0.1308
Surgery duration	0.00	3.42	0.0006
Herniorrhaphy	0.86	4.02	0.0001

*Additional Intermountain risk factor.

E3.5.2. Multivariate analysis of the datasets, including only CABG surgery.

During the univariate analyses of the derivation dataset, 14 different risk factors were included in the model compared to only 7 for the previous total dataset. Each of those 14 risk factors was then included in three different logistic regression analyses using a 60 percent derivation set, a 40 percent validation set and then the combined datasets (Exhibit 56, Exhibit 57, and Exhibit 58). The significance of each of the 14 potential risk factors changed during each test using the three different datasets. For the derivation dataset, nine of the 14 univariate risk factors remained significant in the model compared to only two in the validation set and four when both derivation and validation sets were combined. In this case, only postoperative admission within 30 days and BMI was significant in all three tests. Three of the eight new risk factors made it into the multivariate analysis but only number of surgeons was significant in the derivation analysis. Compared to the previous combined analyses with the total dataset, the combined analysis with only the Intermountain data only identified postop hemoglobin as an additional risk factor along with BMI, history of MRSA and postop admission which were included in the previous all site analysis. Surgery duration was the only significant risk factor from the previous analysis not included in the Intermountain data only. Thus the significant risk factors from the analysis with only the Intermountain data were very similar to those in the analysis with total dataset. This is not surprising since the Intermountain data contributed over 95 percent of the CABG surgeries in the total dataset. The difference would be attributed to the 78 CABG surgeries from the VA and the inclusion of the eight new risk factors.

Exhibit 56. Results of logistic regression using the derivation dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-10.72	-2.92	0.0035
Elective surgery	-1.88	-2.43	0.0150
BMI	0.12	3.21	0.0013
History of cancer	3.30	2.56	0.0104
Charlson score*	-0.63	-1.88	0.0597
Emergency surgery	-1.31	-1.45	0.1479
History of MRSA	3.02	3.57	0.0004
No. of procedures	-1.31	-1.99	0.0466
No. of surgeons*	1.07	2.14	0.0326
Postop admission	1.57	2.44	0.0149
Postop glucose*	0.01	0.96	0.3370
Postop hemoglobin	-0.33	-1.99	0.0471
Preop albumin	0.98	1.72	0.0863
Preop glucose	0.01	2.02	0.0437
Preop stay	-0.45	-1.36	0.1745

*Additional Intermountain risk factor.

Exhibit 57. Results of logistic regression using the validation dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-3.96	-1.18	0.2365
Elective surgery	-0.21	-0.35	0.7255
BMI	0.09	2.55	0.0109
History of cancer[@]			
Charlson score[*]	-0.33	-1.26	0.2071
Emergency surgery	1.08	1.83	0.0672
History of MRSA	0.91	0.94	0.3456
No. of procedures	-0.22	-0.46	0.6476
No. of surgeons[*]	0.50	0.65	0.5161
Postop admission	1.58	2.75	0.0059
Postop glucose[*]	0.00	-0.44	0.6635
Postop hemoglobin	-0.19	-1.35	0.1782
Preop albumin	-0.28	-0.49	0.6246
Preop glucose[*]	-0.01	-0.84	0.3995
Preop stay	0.12	1.68	0.0923

[@]No patients in this dataset with had the specific risk factor and no probability could be calculated.
[*]Additional Intermountain risk factor.

Exhibit 58. Results of logistic regression using the combined dataset for only CABG surgeries.

Variable	Estimate	tValue	Probability
Intercept	-7.42	-3.28	0.0011
admit_physician??	-0.53	-1.18	0.2365
BMI	0.10	4.19	0.0000
History of cancer	1.60	1.71	0.0875
Charlson score[*]	-0.45	-2.48	0.0132
Emergency surgery	0.14	0.31	0.7555
History of MRSA	1.90	3.37	0.0008
No. of procedures	-0.65	-1.68	0.0920
No. of surgeons[*]	0.68	1.73	0.0831
Postop admission	1.50	3.71	0.0002
Postop glucose[*]	0.00	0.53	0.5975
Postop hemoglobin	-0.24	-2.35	0.0186
Preop albumin	0.28	0.72	0.4716
Preop glucose[*]	0.00	0.44	0.6570
Preop stay	0.10	1.37	0.1712

[*]Additional Intermountain risk factor.

E3.5.3. Multivariate analysis of the datasets including only herniorrhaphy.

During the univariate analyses of this derivation dataset, six different risk factors were included in the model compared to 7 for the previous total dataset. Each of those six risk factors was then included in three different logistic regression analyses using a 60 percent derivation set,

a 40 percent validation set and then the combined datasets (Exhibit 59, Exhibit 60, and Exhibit 61). The significance of each of the six potential risk factors changed during each test using the three different datasets. For the derivation and validation datasets, only postop admission remained significant in the model. In the combined dataset, emergency surgery was significant in addition to postop admission. Only preop glucose from the eight new risk factors made it into the model, but was not found to be significant in any of the three separate analyses. Compared to the previous combined analyses with the total dataset, the combined analysis with only the Intermountain data only identified only emergency surgery in addition to postop admission while postop admission was the only risk factor identified in the previous total dataset. Thus, while the risk factors included in the model were mostly different, only emergency surgery was different in the list of significant risk factors. Since Intermountain only contributed 57 percent of the herniorrhaphy data, this similar result was not due to a dominance of Intermountain data in this case.

Exhibit 59. Results of logistic regression using the derivation dataset for only herniorrhaphy.

Variable	Estimate	tValue	Probability
Intercept	-3.04	-1.16	0.2480
Emergency surgery	0.76	1.01	0.3111
Previous DVT	1.26	1.87	0.0609
Male	0.34	0.60	0.5500
Postop admission	3.17	3.98	0.0001
Preop albumin	-0.75	-1.30	0.1965
Preop glucose*	0.00	-0.44	0.6643

*Additional Intermountain risk factor.

Exhibit 60. Results of logistic regression using the validation dataset for only herniorrhaphy.

Variable	Estimate	tValue	Probability
Intercept	-3.94	-0.92	0.3794
Emergency surgery	1.21	1.83	0.0676
Previous DVT[@]			
Male	-0.87	-1.37	0.1703
Postop admission	2.91	3.67	0.0002
Preop albumin	-0.44	-0.42	0.6821
Preop glucose[*]	0.00	0.29	0.7724

[@]No patients in this dataset with had the specific risk factor and no probability could be calculated.

*Additional Intermountain risk factor.

Exhibit 61. Results of logistic regression using the combined dataset for only herniorrhaphy.

Variable	Estimate	tValue	Probability
Intercept	-4.00	-1.96	0.0561
Emergency surgery	1.10	2.37	0.0178
Previous DVT	0.52	0.88	0.3775
Male	-0.23	-0.55	0.5789
Postop admission	3.08	5.53	0.0000
Preop albumin	-0.43	-0.96	0.3446
Preop glucose[*]	0.00	-0.37	0.7127

[*]Additional Intermountain risk factor.

E3.5.4. Multivariate analysis of the datasets including only total hip surgery.

During the univariate analyses of the derivation dataset, 12 different risk factors were included in the model compared to eight for the previous total dataset. Each of those 12 risk factors was then included in three different logistic regression analyses using a 60 percent derivation set, a 40 percent validation set and then the combined datasets (Exhibit 62, Exhibit 63, and Exhibit 64). The significance of each of the 12 potential risk factors changed during each test using the three different datasets. For the derivation dataset, only three of the 12 univariate risk factors remained significant in the model compared to only one in the validation set and then four when both derivation and validation sets were combined. Only postoperative admission within 30 days was significant in all three tests. Only two, Charlson score and preop glucose, of the eight new risk factors made it into the multivariate analysis but neither was significant in any of the three analyses. Of interest for total hip surgery, although not all of the risk factors identified during the univariate analyses were the same, the same four risk factors were significant in the logistic regression combined analyses using the Intermountain data alone and the previous combined analyses with the total dataset. Total hip surgeries from Intermountain contributed to 85 percent of the total dataset.

Exhibit 62. Results of logistic regression using the derivation dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-3.76	-1.01	0.3183
Age	-0.02	-1.11	0.2674
Chronic lung disease	1.85	2.33	0.0197
Charlson score*	-0.47	-1.07	0.2827
Emergency surgery	1.75	2.38	0.0172
General anesthesia	1.22	1.46	0.1447
Male	1.11	1.85	0.0645
Postop admission	3.55	3.36	0.0008
Preop glucose*	-0.01	-0.46	0.6616
Preop hematocrit	-0.03	-0.28	0.7829
Preop hemoglobin	-0.25	-0.68	0.4966
Surgery duration	0.01	1.91	0.0571
Wound class	0.54	1.38	0.1681

*Additional Intermountain risk factor.

Exhibit 63. Results of logistic regression using the validation dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-3.87	-1.34	0.1815
Age	-0.01	-0.62	0.5327
Chronic lung disease	0.44	0.67	0.4999
Charlson score*	0.08	0.33	0.7412
Emergency surgery	0.95	1.18	0.2372
General anesthesia	1.12	1.44	0.1489
Male	-0.78	-1.33	0.1842
Postop admission	1.37	2.63	0.0086
Preop glucose*	-0.01	-1.22	0.2331
Preop hematocrit	-0.06	-0.66	0.5089
Preop hemoglobin	0.13	0.46	0.6480
Surgery duration	0.01	1.75	0.0801
Wound class	0.26	0.49	0.6263

*Additional Intermountain risk factor.

Exhibit 64. Results of logistic regression using the combined dataset for only total hip surgery.

Variable	Estimate	tValue	Probability
Intercept	-3.87	-1.68	0.1005
Age	-0.02	-1.23	0.2195
Chronic lung disease	1.07	2.32	0.0205
Charlson score*	-0.14	-0.63	0.5271
Emergency surgery	1.31	2.60	0.0094
General anesthesia	1.03	1.87	0.0618
Male	0.11	0.28	0.7768
Postop admission	2.04	4.70	0.0000
Preop glucose*	-0.01	-0.96	0.3454
Preop hematocrit	-0.05	-0.78	0.4367
Preop hemoglobin	-0.01	-0.03	0.9787
Surgery duration	0.01	2.94	0.0033
Wound class	0.34	1.13	0.2605

*Additional Intermountain risk factor.

E3.5.5. Multivariate analysis of the datasets including only total knee surgery.

During the univariate analyses of this derivation dataset, eight different risk factors were included in the model compared to 5 for the previous total dataset. Each of those eight risk factors was then included in three different logistic regression analyses using a 60 percent derivation set, a 40 percent validation set and then the combined datasets (Exhibit 65, Exhibit 66, and Exhibit 67). The significance of each of the six potential risk factors changed during each test using the three different datasets. For the derivation dataset, Charlson score, male, history of MRSA and postop admission remained significant, only history of MRSA and postop admission were significant in the validation dataset and male was included along with history of MRSA and postop admission in the combined dataset. While it was included in many of the other Intermountain analyses, this was the first time Charlson score was found to be significant. Compared to the previous combined analyses with the total dataset, the combined analysis with only the Intermountain data only identified of the five risk factors identified in the previous total dataset. Number of procedures and preop hematocrit were additionally found significant in the total dataset. The Intermountain total knee data contributed to 90 percent of the total dataset.

Exhibit 65. Results of logistic regression using the derivation dataset for only total knee surgery.

Variable	Estimate	tValue	Probability
Intercept	-6.27	-3.13	0.0025
Charlson score*	0.32	2.03	0.0423
General anesthesia	-0.48	-1.22	0.2231
Male	0.90	2.20	0.0280
History of MRSA	1.49	2.69	0.0071
Postop admission	2.97	4.83	0.0000
Postop hematocrit	0.06	0.86	0.3922
Preop hematocrit	0.09	0.57	0.5742
Preop hemoglob	-0.50	-1.09	0.2835
Surgery duration	0.00	1.25	0.2104

*Additional Intermountain risk factor.

Exhibit 66. Results of logistic regression using the validation dataset for only total knee surgery.

Variable	Estimate	tValue	Probability
Intercept	-3.76	-2.23	0.0256
Charlson score*	-0.51	-1.75	0.0793
General anesthesia	0.53	1.33	0.1819
Male	0.60	1.70	0.0896
History of MRSA	2.12	4.45	0.0000
Postop admission	1.74	4.66	0.0000
Postop hematocrit	-0.01	-0.10	0.9173
Preop hematocrit	0.06	0.54	0.5918
Preop hemoglob	-0.37	-1.11	0.2670
Surgery duration	0.01	1.64	0.1007

*Additional Intermountain risk factor.

Exhibit 67. Results of logistic regression using the combined dataset for only total knee surgery.

Variable	Estimate	tValue	Probability
Intercept	-4.57	-3.79	0.0002
Charlson score*	0.03	0.19	0.8475
General anesthesia	0.11	0.43	0.6692
Male	0.73	2.75	0.0059
History of MRSA	1.76	5.01	0.0000
Postop admission	2.15	6.99	0.0000
Postop hematocrit	0.02	0.46	0.6493
Preop hematocrit	0.04	0.37	0.7176
Preop hemoglob	-0.33	-1.08	0.2875
Surgery duration	0.00	1.78	0.0759

*Additional Intermountain risk factor.

E3.5.6. Multivariate analysis of the dataset including only appendectomy surgery at Intermountain healthcare

Although all the appendectomy data in the total dataset was from Intermountain, we analyzed the appendectomy data again with the eight new potential risk factors included. During the univariate analyses of the derivation dataset, 10 different risk factors were included in the model compared to only seven for the previous dataset. Each of those 10 risk factors was then included again in three different logistic regression analyses using a 60 percent derivation set, a 40 percent validation set and then the combined datasets (Exhibit 68, Exhibit 69, and Exhibit 70). The significance of each of the 10 potential risk factors changed during each test using the three different datasets. For the derivation dataset, only three of the 10 univariate risk factors remained significant in the model compared to three in the validation set and three when both derivation and validation sets were combined. Postop hematocrit along with postoperative admission within 30 days was significant in all three tests. While Charlson score and preop glucose were the only two of the eight new risk factors that made it into the multivariate analysis, only Charlson score was significant in the derivation analysis. The inclusion of the eight new potential risk factors in this analysis did impact the list of univariate risk factors included in the logistic analysis and the significance of each. Postop hemoglobin was significant in the analysis that included the new risk factors in addition to the same other two, postop admission and postop hematocrit, that were significant in the previous analysis without them.

Exhibit 68. Results of logistic regression using the derivation dataset for only appendectomy surgeries.

Variable	Estimate	tValue	Probability
Intercept	-2.31	-0.88	0.3806
Age	0.01	1.12	0.2634
Charlson score*	0.20	1.96	0.0496
Diabetes	-1.07	-1.12	0.2646
General anesthesia	-1.25	-0.82	0.4100
History of MRSA	1.51	1.32	0.1860
Postop admission	2.21	4.56	0.0000
Postop hematocrit	-0.14	-2.00	0.0478
Postop hemoglobin	0.27	1.02	0.3178
Preop glucose*	0.01	0.99	0.3222
Preop hemoglobin	-0.16	-0.68	0.5045

*Additional Intermountain risk factor.

Exhibit 69. Results of logistic regression using the validation dataset for only appendectomy surgeries.

Variable	Estimate	tValue	Probability
Intercept	-4.95	-1.27	0.2050
Age	-0.01	-0.21	0.8311
Charlson score*	-0.32	-0.78	0.4367
Diabetes	1.90	1.24	0.2134
General anesthesia@			
History of MRSA@			
Postop admission	2.40	2.92	0.0035
Postop hematocrit	-0.28	-2.09	0.0392
Postop hemoglobin	0.84	2.06	0.0434
Preop glucose*	-0.01	-0.92	0.3632
Preop hemoglobin	-0.09	-0.31	0.7564

@No patients in this dataset with had the specific risk factor and no probability could be calculated.
*Additional Intermountain risk factor.

Exhibit 70. Results of logistic regression using the combined dataset for only appendectomy surgeries.

Variable	Estimate	tValue	Probability
Intercept	-3.34	-1.48	0.1399
Age	0.01	1.01	0.3142
Charlson score*	0.16	1.65	0.0995
Diabetes	-0.62	-0.79	0.4299
General anesthesia	-1.72	-1.33	0.1823
History of MRSA	0.97	0.89	0.3710
Postop admission	2.27	5.43	0.0000
Postop hematocrit	-0.13	-2.24	0.0263
Postop hemoglobin	0.40	2.15	0.0317
Preop glucose*	0.00	0.65	0.5150
Preop hemoglobin	-0.16	-0.97	0.3379

*Additional Intermountain risk factor.

Chapter 6. Conclusions and Recommendations

Challenges Encountered, and Strategies for Overcoming Them

Below we note specific challenges encountered in executing our work plan as planned and those strategies used to overcome these.

November 2009

Task 2: The VA system relies on CPT codes for administrative coding whereas other systems use ICD-9 codes.

A mapping of ICD-9/CPT codes for the specific procedures being evaluated in the project was done to allow for a fully representative sample of surgeries at the VA system.

February 2010

Task 2 & 3: Craig Gale left Intermountain

Jef Huntington, from Intermountain, joined the project team for the analytic work of tasks 2 and 3.

March 2010

Task 3: Russ Staheli left Intermountain

Jef Huntington was included to coordinate the data pulls for task 3.

April 2010

Task 3: After completing the master risk factor list, many identified variables were dependent upon the definition of the variable, e.g. chronic diseases had to be clearly defined.

An SSI risk comorbidity table was created using the ICD-9 /CPT map along with other identified factors ICD-9 codes to standardize the definitions.

June 2010

Task 3: Certain identified SSI risk factors were reevaluated for the data collection process, e.g. Anemia was identified as a risk factor and was to be recorded as a yes/no variable.

Instead of initially defining conditions and providing a yes/no value, it was decided that measured values would be more useful. In the case of anemia, hemoglobin levels were to be recorded—from which a set definition of Anemia could be derived.

July 2010

Task 2: While algorithm development and testing was planned to use SLC VAMC and Intermountain NSPQIP data, the data was only available at SLC VAMC.

It was decided that National VA NSQIP data could be used to develop and test the algorithm, then validated at the other systems.

Task 3: The risk factor data collection at each site was delayed due to continued refining of the master risk factor list.

As data collection started, Intermountain did a provisional analysis of about 20,000 patients to determine collection reliability. All sites recorded the data collection process and noted any elements that were unreliable or difficult to obtain. The master risk factor list was updated as necessary to include reliable and obtainable elements.

October 2010

Task 1: A no-cost extension was considered to use funds for travel and conference attendance to present the work of the project after the project end date.

The no cost extension was denied. However, the possibility of paying for expenses associated with the dissemination plan prior to the project end date was considered as a possible option and required further review from AHRQ Contract Officers.

Task 3: As the data pulls progressed across the systems it was realized that the data sources varied between each system.

In order to aid the implementation of the tool at other systems, it was decided to record, with detailed specificity, where the data were found at each organization.

Task 4: Given the delay of conducting the nursing focus group as originally planned, a repurposed focus group strategy was proposed.

Repurposed focus groups were submitted for approval and approved to conduct 2 separate groups to solicit input from key stakeholders on adoption and implementation as well as developing use cases for the e-detection surveillance tool.

January 2011

Task 2: With IRB/Privacy Board requirements, data could not be sent to SLC VAMC for validation of the algorithm.

The algorithm was sent to each organization to test and validate with their data where chart reviews will be performed on all positives produced by the algorithm.

Lessons Learned and Recommendations for Next Steps

There are several lessons to be learned from this work, summarized in the following bullet points:

Task 1

- In-person meetings among project team members, held in addition to regularly scheduled teleconferences, confirmed our expectations of their value for promoting teambuilding and collaboration among geographically distributed team members.

Task 2:

- The most appropriate use of automated systems, whether alone or in combination with manual surveillance, will take careful consideration of the purpose and requirements of the events being surveilled. The performance of automated systems may vary, particularly when attempting to detect events that occur in the outpatient setting where differences in data availability may be pronounced. More work is necessary to improve the discriminability index of electronic algorithms, but allowing IP to select rules that suit their own needs may be a reasonable measure in the interim.
- Current estimates for national rates of SSI are unknown. Our smaller group of hospitals is only informative in a very limited way. More accurate estimates would require sophisticated patient case mix adjustment and a much larger sampling of hospitals.
- There may be value in exploring natural language programming (NLP) and what could be added from text notes. A new study shows that NLP will be more beneficial for the electronic identification of hospital-acquired wounds than bacteremias, UTIs, respiratory infections, etc. (publication in progress).

Task 3:

- Postdischarge surveillance remains a challenge, requiring data from the full continuum of care (inpatient and outpatient). Postop admission within 30 days is a common "trigger" used by many facilities that rely on manual SSI surveillance for postdischarge infection identification. Integrated inpatient/outpatient medical records are expected to have more utility for electronic algorithms.
- No single set of risk factors that can be used to predict SSI across all types of surgical procedures or facilities. This study found that SSI risk factors are dependent on the type of surgical procedure. Thus, SSI rate comparison needs to be at the surgical procedure level and not the surgical service level, i.e., orthopedics, general surgery, thoracic, etc. In addition, SSI rates should also be compared at the facility level against its own baseline rates.

Task 4:

- An extreme burden of unsupported practitioner SSI surveillance exists.
- There is documented need for enhanced risk factor assessment (surgeon focus group) and receptivity toward cognitive support from an electronic surveillance tool (nurse focus groups).

Task 5:

- An analysis of DH publicly reported data revealed SIR was superior than the NHSN basic risk index in predicting SSI risk. The SIR uses logistic regression modeling and takes into account more variables and procedure-specific risk factors. We hypothesize that improved risk adjustment is due to consideration of these extra risk factors. Although superior, the SIR still is not broadly applicable to all procedures and settings. Based on this, we are now looking at surgery specific risk factor assessment,

accounting for differential impact of, e.g., smoking on infection risk for hernia vs. CABG, etc. We need to investigate risk factors as surgery specific. Now with more publicly reported data we can use that to get the needed large numbers.

Several recommendations for next steps have emerged from our work. These are:

1. Algorithms should be validated at each system to which they will be introduced, in a manner akin to the quality assurance policies regarding new laboratory equipment.
2. Algorithms will need to be trained and validated on even broader scales that demonstrate more variation in clinical practice and electronic systems.
3. Validate risk factors more broadly on a national scale.
4. We should investigate ways to develop electronic algorithms that might search multiple inpatient and/or outpatient networks to help ascertain postdischarge SSIs.
5. We need further validation of SIR on publicly reported data; perhaps validate on subgroups of settings (public health safety net, academic, community, etc.).
6. Explore how risk factors may be more or less relevant to specific procedures (e.g., risk of smoking on herniorraphies vs. CABG.
7. National estimates of SSI should not be pursued without larger datasets that are representative of the variation among the nation's hospitals.
8. We should explore natural language processing (NLP) methodology to extract more information from text notes. For instance, a recent study showed that NLP could identify a number of postoperative surgical complications in the Veterans Health Administration[77].
9. We should consider exploration of more sophisticated decision support methods that deliver the probability of SSI and/or important nuance information instead of binary yes/no information (which loses much of the original information content).

References

1. Goldrick B, Larson E. Assessment of infection control programs in Maryland skilled-nursing long-term care facilities. Am J Infect Control. Apr 1994;22(2):83-89.
2. Goldrick BA. The Certification Board of Infection Control and Epidemiology white paper: the value of certification for infection control professionals. Am J Infect Control. Apr 2007;35(3):150-156.
3. Stevenson KB, Murphy CL, Samore MH, et al. Assessing the status of infection control programs in small rural hospitals in the western United States. Am J Infect Control. Aug 2004;32(5):255-261.
4. Rubin MA, Mayer J, Greene T, et al. An agent-based model for evaluating surveillance methods for catheter-related bloodstream infection. AMIA Annu Symp Proc. 2008:631-635.
5. Brossette SE, Hacek DM, Gavin PJ, et al. A laboratory-based, hospital-wide, electronic marker for nosocomial infection: the future of infection control surveillance? Am J Clin Pathol. Jan 2006;125(1):34-39.
6. Gaynes RP, Culver DH, Horan TC, et al. Surgical site infection (SSI) rates in the United States, 1992-1998: the National Nosocomial Infections Surveillance System basic SSI risk index. Clin Infect Dis. Sep 1 2001;33 Suppl 2:S69-77.
7. Edmond MB, White-Russell MB, Ober J, et al. A statewide survey of nosocomial infection surveillance in acute care hospitals. Am J Infect Control. Oct 2005;33(8):480-482.
8. Rowell KS, Turrentine FE, Hutter MM, et al. Use of national surgical quality improvement program data as a catalyst for quality improvement. J Am Coll Surg. Jun 2007;204(6):1293-1300.
9. Verkkala K, Valtonen V, Jarvinen A, et al. Fever, leucocytosis and C-reactive protein after open-heart surgery and their value in the diagnosis of postoperative infections. Thorac Cardiovasc Surg. Apr 1987;35(2):78-82.
10. de Zwart PM, Muller JE. [C-reactive protein as a postoperative parameter of infection in hip joint replacement and knee joint operations]. Aktuelle Traumatol. Aug 1994;24(5):184-187.
11. Yokoe DS, Platt R. Surveillance for surgical site infections: the uses of antibiotic exposure. Infect Control Hosp Epidemiol. Nov 1994;15(11):717-723.
12. Bourguignat A, Ferard G, Jenny JY, et al. Diagnostic value of C-reactive protein and transthyretin in bone infections of the lower limb. Clin Chim Acta. Nov 15 1996;255(1):27-38.
13. Rosas MH, Leclercq S, Pegoix M, et al. Contribution of laboratory tests, scintigraphy, and histology to the diagnosis of lower limb joint replacement infection. Rev Rhum Engl Ed. Jul-Sep 1998;65(7-9):477-482.
14. Jenny JY, Gaudias J, Bourguignat A, et al. [C-reactive protein and transthyretin in early diagnosis of infection after open fractures of the lower limbs (a preliminary study)]. Rev Chir Orthop Reparatrice Appar Mot. Jul 1999;85(4):321-327.
15. Rothenburger M, Markewitz A, Lenz T, et al. Detection of acute phase response and infection. The role of procalcitonin and C-reactive protein. Clin Chem Lab Med. Mar 1999;37(3):275-279.
16. Sands K, Vineyard G, Livingston J, et al. Efficient identification of postdischarge surgical site infections: use of automated pharmacy dispensing information, administrative data, and medical record information. J Infect Dis. Feb 1999;179(2):434-441.
17. Bitkover CY, Hansson LO, Valen G, et al. Effects of cardiac surgery on some clinically used inflammation markers and procalcitonin. Scand Cardiovasc J. Jun 2000;34(3):307-314.
18. Platt R, Yokoe DS, Sands KE. Automated methods for surveillance of surgical site infections. Emerg Infect Dis. Mar-Apr 2001;7(2):212-216.
19. Lannergard A, Hersio K, Larsson A, et al. Evaluation of laboratory markers for the detection of infections in open-heart surgery patients. Scand J Infect Dis. 2003;35(2):121-126.
20. Miner AL, Sands KE, Yokoe DS, et al. Enhanced identification of postoperative infections among outpatients. Emerg Infect Dis. Nov 2004;10(11):1931-1937.

21. Yokoe DS, Noskin GA, Cunnigham SM, et al. Enhanced identification of postoperative infections among inpatients. Emerg Infect Dis. Nov 2004;10(11):1924-1930.

22. Codine P, Barbotte E, Denis-Laroque F, et al. [C-reactive protein, leukocyte count and D-dimer monitoring after orthopedic surgery: early diagnosis of infectious or thromboembolic complications. Part one: C-reactive protein and leukocyte count as an aid in diagnosing postoperative infection]. Ann Readapt Med Phys. Nov 2005;48(8):590-597.

23. Falcoz PE, Laluc F, Toubin MM, et al. Usefulness of procalcitonin in the early detection of infection after thoracic surgery. Eur J Cardiothorac Surg. Jun 2005;27(6):1074-1078.

24. Huenger F, Schmachtenberg A, Haefner H, et al. Evaluation of postdischarge surveillance of surgical site infections after total hip and knee arthroplasty. Am J Infect Control. Oct 2005;33(8):455-462.

25. Michelson J. Improved detection of orthopaedic surgical site infections occurring in outpatients. Clin Orthop Relat Res. Apr 2005(433):218-224.

26. Spolaore P, Pellizzer G, Fedeli U, et al. Linkage of microbiology reports and hospital discharge diagnoses for surveillance of surgical site infections. J Hosp Infect. Aug 2005;60(4):317-320.

27. Petherick ES, Dalton JE, Moore PJ, et al. Methods for identifying surgical wound infection after discharge from hospital: a systematic review. BMC Infect Dis. 2006;6:170.

28. Takahashi J, Shono Y, Hirabayashi H, et al. Usefulness of white blood cell differential for early diagnosis of surgical wound infection following spinal instrumentation surgery. Spine (Phila Pa 1976). Apr 20 2006;31(9):1020-1025.

29. Yasmin D, Bulut G, Yildiz M. [Can procalcitonin be used for the diagnosis and follow-up of postoperative complications after fracture surgery?]. Acta Orthop Traumatol Turc. 2006;40(1):15-21.

30. Huang SS, Livingston JM, Rawson NS, et al. Developing algorithms for healthcare insurers to systematically monitor surgical site infection rates. BMC Med Res Methodol. 2007;7:20.

31. Dupont C, Rodenbach J, Flachaire E. The value of C-reactive protein for postoperative monitoring of lower limb arthroplasty. Ann Readapt Med Phys. Jun 2008;51(5):348-357.

32. Moyad TF, Thornhill T, Estok D. Evaluation and management of the infected total hip and knee. Orthopedics. Jun 2008;31(6):581-588; quiz 589-590.

33. Schinsky MF, Della Valle CJ, Sporer SM, Paprosky WG. Perioperative testing for joint infection in patients undergoing revision total hip arthroplasty. J Bone Joint Surg Am. Sep 2008;90(9):1869-1875.

34. Stevenson KB, Khan Y, Dickman J, et al. Administrative coding data, compared with CDC/NHSN criteria, are poor indicators of health care-associated infections. Am J Infect Control. Apr 2008;36(3):155-164.

35. Bolon MK, Hooper D, Stevenson KB, et al. Improved surveillance for surgical site infections after orthopedic implantation procedures: extending applications for automated data. Clin Infect Dis. May 1 2009;48(9):1223-1229.

36. Herold C, Jokuszies A, Steiert A, et al. C-reactive protein is not a reliable marker of infection in flap reconstructive procedures of lower extremities. J Plast Reconstr Aesthet Surg. Dec 2009;62(12):e670-671.

37. Kang BU, Lee SH, Ahn Y, Choi WC, Choi YG. Surgical site infection in spinal surgery: detection and management based on serial C-reactive protein measurements. J Neurosurg Spine. Aug 2010;13(2):158-164.

38. Leth RA, Moller JK. Surveillance of hospital-acquired infections based on electronic hospital registries. J Hosp Infect. Jan 2006;62(1):71-79.

39. Gastmeier P, Brauer H, Hauer T, et al. How many nosocomial infections are missed if identification is restricted to patients with either microbiology reports or antibiotic administration? Infect Control Hosp Epidemiol. Feb 1999;20(2):124-127.

40. Hirschhorn LR, Currier JS, Platt R. Electronic surveillance of antibiotic exposure and coded discharge diagnoses as indicators of postoperative infection and other quality assurance measures. Infect Control Hosp Epidemiol. Jan 1993;14(1):21-28.

41. Platt R, Kleinman K, Thompson K, et al. Using automated health plan data to assess infection risk from coronary artery bypass surgery. Emerg Infect Dis. Dec 2002;8(12):1433-1441.

42. Yokoe DS, Shapiro M, Simchen E, et al. Use of antibiotic exposure to detect postoperative infections. Infect Control Hosp Epidemiol. May 1998;19(5):317-322.

43. Ekman G, Weitoft GR, Nyren O, et al. National Surveillance of Surgical-Site Infection through Register-based Analysis of Antibiotic Use after Inguinal Hernia Repair. British Journal of Surgery. 2010;97:1722-1729.

44. Sands KE, Yokoe DS, Hooper DC, et al. Detection of postoperative surgical-site infections: comparison of health plan-based surveillance with hospital-based programs. Infect Control Hosp Epidemiol. Oct 2003;24(10):741-743.

45. Gonzalez C, Best B, Healy AF, et al. A cognitive modeling account of simultaneous learning and fatigue effects. Cognitive Systems Research. March 2011 2011;12(1):19-32.

46. Jones MM SM, Nechodom P, Rubin MA. Surveillance for hospital-acquired VRE and MRPA infections using electronic algorithms [Abstract]. Accepted as a poster. Received IDSA fellow's travel grant. Paper presented at: Infectious Disease Society of America2009; Philadelphia, Pennsylvania.

47. Jones MM SM, Jernigan JA, O'Connor-Wright S, et al. Surveillance for hospital-acquired methicillin-resistant Staphylococcus aureus(MRSA) infections using electronic algorithm [Abstract]. Accepted and oral presentation delivered by Makoto Jones 3/21/09. Paper presented at: SHEA2009; San Diego, California.

48. Jones MM SM, Nechodom P, Rubin MA. Comparison of Automated and Manual Methods of Estimating Methicillin-Resistant Staphylococcus aureus (MRSA) Healthcare-Associated Infection (HAI) Rates [Abstract]. Accepted as a poster. Paper presented at: SHEA2010; Atlanta, Georgia.

49. Moreno LF, Stratton HH, Newell JC, et al. Mathematical coupling of data: correction of a common error for linear calculations. J Appl Physiol. Jan 1986;60(1):335-343.

50. Cardo DM, Falk PS, Mayhall CG. Validation of surgical wound surveillance. Infect Control Hosp Epidemiol. Apr 1993;14(4):211-215.

51. Borlawsky T, Hota B, Lin MY, et al. Development of a reference information model and knowledgebase for electronic bloodstream infection detection. AMIA Annu Symp Proc. 2008:56-60.

52. Quaglini S. Information models for data sharing. Neurol Sci. Jun 2006;27 Suppl 3:S281-283.

53. Klevens RM, Edwards JR, Richards CL, Jr., et al. Estimating health care-associated infections and deaths in U.S. hospitals, 2002. Public Health Rep. Mar-Apr 2007;122(2):160-166.

54. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. Am J Infect Control. Dec 2004;32(8):470-485.

55. Leekha S, Sampathkumar P, Berry DJ, et al. Should national standards for reporting surgical site infections distinguish between primary and revision orthopedic surgeries? Infect Control Hosp Epidemiol. May 2010;31(5):503-508.

56. Muilwijk J, Walenkamp GH, Voss A, et al. Random effect modelling of patient-related risk factors in orthopaedic procedures: results from the Dutch nosocomial infection surveillance network 'PREZIES'. J Hosp Infect. Mar 2006;62(3):319-326.

57. von Dolinger EJ, de Souza GM, de Melo GB, et al. Surgical site infections in primary total hip and knee replacement surgeries, hemiarthroplasties, and osteosyntheses at a Brazilian university hospital. Am J Infect Control. Apr 2010;38(3):246-248.

58. Wilson J, Ramboer I, Suetens C. Hospitals in Europe Link for Infection Control through Surveillance (HELICS). Inter-country comparison of rates of surgical site infection--opportunities and limitations. J Hosp Infect. Jun 2007;65 Suppl 2:165-170.

59. Kurtz SM, Ong KL, Lau E, et al. Prosthetic joint infection risk after TKA in the Medicare population. Clin Orthop Relat Res. Jan 2010;468(1):52-56.

60. Abboud CS, Wey SB, Baltar VT. Risk factors for mediastinitis after cardiac surgery. Ann Thorac Surg. Feb 2004;77(2):676-683.

61. Filsoufi F, Castillo JG, Rahmanian PB, et al. Epidemiology of deep sternal wound infection in cardiac surgery. J Cardiothorac Vasc Anesth. Aug 2009;23(4):488-494.

62. Mannien J, van der Zeeuw AE, Wille JC, et al. Validation of surgical site infection surveillance in the Netherlands. Infect Control Hosp Epidemiol. Jan 2007;28(1):36-41.

63. Rosenthal R, Weber WP, Marti WR, et al. Surveillance of surgical site infections by surgeons: biased underreporting or useful epidemiological data? J Hosp Infect. Jul 2010;75(3):178-182.

64. Gastmeier P, Kampf G, Hauer T, et al. Experience with two validation methods in a prevalence survey on nosocomial infections. Infect Control Hosp Epidemiol. Sep 1998;19(9):668-673.

65. Mayer J, Howell, J,Green T, et al. Assessing Inter-Rater Reliability (IRR) of Surveillance Decisions by Infection Preventionists (IPs). Atlanta, Georgia: Society for Hospital Epidemiology of America; 2010.

66. Trick WE, Zagorski BM, Tokars JI, et al. Computer algorithms to detect bloodstream infections. Emerg Infect Dis. Sep 2004;10(9):1612-1620.

67. Allami MK, Jamil W, Fourie B, et al. Superficial incisional infection in arthroplasty of the lower limb. Interobserver reliability of the current diagnostic criteria. J Bone Joint Surg Br. Sep 2005;87(9):1267-1271.

68. Leal J, Laupland KB. Validity of electronic surveillance systems: a systematic review. J Hosp Infect. Jul 2008;69(3):220-229.

69. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA. Aug 24 2011;306(8):848-855.

70. Ferucci D, Brown E, Chu-Carroll J, et al. Buidling Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010 2010:59-79.

71. Zou H Y, Y. Combining time series models for forecasting. International Journal of Forecasting. 2004;20:69-84.

72. Cherkassky V, Ma Y. Multiple model regression estimation. IEEE Trans Neural Netw. Jul 2005;16(4):785-798.

73. Huntington JL, Dueck A. Handling missing data. Curr Probl Cancer. 2005;29(6).

74. Wood AM, White IR, Thompson SG. Are missing outcome data adequately handled? A review of published randomized controlled trials in major medical journals. clin Trials. 2004;1.

75. Chapter 9. The MI Procedure. In: SAS OnlineDoc™ Version 8. Cary, NC: SAS Institute. http://support.sas.com/rnd/app/papers/miv802.pdf

76. Inacio MCS, Paxton EW, Chen Y, , et al. Leveraging electronic medical records for surveillance of surgical site infection in a total joint replacement population. Infection Control and Hospital Epidemiology 2011;32(4):351-359.

77. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA : the journal of the American Medical Association. Aug 24 2011;306(8):848-855.

U.S. Department of Health and Human Services
Agency for Healthcare Research and Quality
540 Gaither Road
Rockville, MD 20850

AHRQ Pub. No. 12-0046-EF
August 2012