Automated Methods for Surveillance of Surgical Site Infections

Richard Platt,*† Deborah S. Yokoe,† Kenneth E. Sands,‡ and the CDC Eastern Massachusetts Prevention Epicenter Investigators1

*Harvard Medical School and Harvard Pilgrim Health Care, Boston, Massachusetts, USA; †Harvard Medical School and Brigham and Women’s Hospital, Boston, Massachusetts, USA; and ‡Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Automated data, especially from pharmacy and administrative claims, are available for much of the U.S. population and might substantially improve both inpatient and postdischarge surveillance for surgical site infections complicating selected procedures, while reducing the resources required. Potential improvements include better sensitivity, less susceptibility to interobserver variation, more uniform availability of data, more precise estimates of infection rates, and better adjustment for patients’ coexisting illness.

The Centers for Disease Control and Prevention (CDC) recommends routine surveillance for surgical site infections (1); accrediting agencies such as the Joint Commission for Accreditation of Healthcare Organizations require it. Surveillance identifies clusters of infection, establishes baseline risks for infection, provides comparisons between institutions or surgical specialties, identifies risk factors, and permits evaluation of control measures (2). Achieving these goals requires health-care systems to have access to different information types (Table 1).

An ideal surveillance system should have several attributes, including meaningful definitions of infection, consistent interpretation of classification criteria, applicability to procedures performed in both inpatient and ambulatory facilities, ability to detect events after discharge, sufficient precision to distinguish small absolute differences in attack rates, ability to adjust for different distribution of severity of illness across populations, and reasonable cost. Most current systems lack at least one of these attributes; for example, the system recommended by CDC’s Hospital Infection Control Practices Advisory Committee (HICPAC) (3) is excellent for clinical decision-making, but some elements are difficult to apply for surveillance purposes. Information required to apply some of its criteria may not be available for all cases; for example, the criterion of recovery of microbial growth from a normally sterile site may be affected by variation in obtaining specimens for culture. Some elements of CDC’s National Nosocomial Infections (NNIS) System definition require substantial judgment or interpretation. An example is determining whether purulent drainage is present. An attending physician’s diagnosis is sufficient, although the way physicians record or confirm their diagnoses may differ.

Table 1. Goals and needs of surgical site infection surveillance (2)

<table>
<thead>
<tr>
<th>Goal</th>
<th>Principal needs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control of clusters</td>
<td>Real-time detection of events.</td>
</tr>
<tr>
<td>Identify clusters of infection</td>
<td>Attack rates and case-mix adjustment are not a high priority. Should include all patients.</td>
</tr>
<tr>
<td>Establish baseline infection rates</td>
<td>Sufficient precision to identify absolute differences of a few percent. Typically includes all patients.</td>
</tr>
<tr>
<td>Comparison of institutions or surgical specialities</td>
<td>Identical detection methods that are applied and interpreted identically across sites. Sufficient precision. Comparably ascertained rates of over time.</td>
</tr>
<tr>
<td>Evaluate control measures</td>
<td>(in the usual situation no randomized trial).</td>
</tr>
<tr>
<td>Research on epidemiology of infection</td>
<td>Detailed data on many attributes of patients and procedures. Population can be small, but must be representative.</td>
</tr>
</tbody>
</table>

For these reasons, case ascertainment is affected by considerable interobserver variability (4). Although most surgical site infections become manifest after the patient is discharged from the hospital (5-12), there is no accepted method for detecting them (13). The most widely described method of conducting postdischarge surveillance is questionnaire reporting by surgeons. This method has been shown to have poor sensitivity (15%) and...
positive predictive value (28%), even when surgeons are compliant in returning the questionnaires (5). Moreover, a questionnaire-based surveillance system requires substantial resources. Reporting by patients via questionnaires also has poor sensitivity (28%) because many patients do not return questionnaires mailed to them a month after surgery. Telephone questionnaires have been used effectively but are too resource intensive for routine use.

Many procedures must be monitored to allow confident conclusions that relatively small differences in observed attack rates do not reflect chance variations. Identifying these small differences, understanding their cause, and undertaking quality improvement programs to reduce their occurrence would have large consequences when applied to the >45 million surgical procedures performed annually in the United States (14). Reducing the overall infection rate by a quarter of a percent would prevent >100,000 infections per year. For coronary artery bypass surgery alone, a one percentage point decrease in the risk for infection would prevent >3,500 infections per year in the United States (15). Because of the need to observe large numbers of procedures, conducting surveillance for the entire surgical population is desirable. However, to conserve scarce resources, some programs survey only a fraction of their procedures or rotate surveillance among different procedure types.

Determining whether relatively small differences in infection rates result from differences in care rather than in patients’ susceptibility to infection requires robust risk-adjustment methods that can take into account different case-mixes in different institutions. Available methods do not have optimal resolution and depend in part on the Anesthesia Society of America (ASA) score (3,16). The ASA score, a subjective assessment of the patient’s overall health status, may reflect interobserver variability (17) that can adversely affect stratification of risk for surgical infection (18).

Automated methods to augment current surveillance methods should improve the quality of surveillance for surgical site infections and reduce the resources required. To achieve these goals, surveillance should be based on the growing body of data that health-care systems, including hospitals, physicians’ offices, health maintenance organizations (HMOs), and insurance companies, routinely collect during care delivery. Many types of automated data are now or will soon become widely available, including information about patients, surgical procedures, and patients’ postoperative courses (Table 2). Three ways to use these data to support surveillance programs are inpatient surveillance, postdischarge surveillance, and case-mix adjustment.

### Inpatient Surveillance for Surgical Site Infections

One of the most widely available types of automated data useful for inpatient surveillance is antibiotic exposure data from pharmacy dispensing records. Studies have indicated that antibiotic exposure is a sensitive indicator of infection (19,20), since relatively few serious infections are managed without antibiotics. Poor specificity (too many false positives) has been a major problem, however, because antibiotics are so widely used after surgery for extended prophylaxis, empiric therapy of suspected infection, and treatment of infections other than surgical site infections.

One way to improve the usefulness of postoperative antibiotic exposure as a marker of infection is to consider the timing and duration of administration, rather than just its

<table>
<thead>
<tr>
<th>Type of information</th>
<th>Availability of this information in specific locations</th>
<th>Automated medical records in physicians’ offices</th>
<th>Payors (HMOs, insurers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic/personal info</td>
<td>Hospitals Usually</td>
<td>Usually</td>
<td>Usually</td>
</tr>
<tr>
<td>Age</td>
<td>Usually</td>
<td>Usually</td>
<td>Usually</td>
</tr>
<tr>
<td>Smoking status</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Body mass index</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Preoperative health status</td>
<td>Diagnoses Sometimes</td>
<td>Usually</td>
<td>Usually</td>
</tr>
<tr>
<td>Procedures</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Drug therapy</td>
<td>Sometimes</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>ASA score</td>
<td>Sometimes</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>Procedure data</td>
<td>Type (ICD-9, CPT) Usually</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Duration</td>
<td>Sometimes</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>Inpatient postoperative care</td>
<td>Diagnoses Usually</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Reoperation</td>
<td>Usually</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>Incision and drainage</td>
<td>Usually</td>
<td>Rarely</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Microbiology data</td>
<td>Usually</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>Antibiotic therapy</td>
<td>Usually</td>
<td>Rarely</td>
<td>Usually</td>
</tr>
<tr>
<td>Postdischarge care</td>
<td>Diagnoses Rarely</td>
<td>Usually</td>
<td>Usually</td>
</tr>
<tr>
<td>Reoperation in another hospital</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
<tr>
<td>Incision and drainage</td>
<td>Rarely</td>
<td>Usually</td>
<td>Usually</td>
</tr>
<tr>
<td>Microbiology data</td>
<td>Rarely</td>
<td>Usually</td>
<td>Sometimes</td>
</tr>
<tr>
<td>Antibiotic therapy</td>
<td>Rarely</td>
<td>Sometimes</td>
<td>Rarely</td>
</tr>
</tbody>
</table>

*Excludes hospital-based physicians’ offices.

occurrence. Quantitative antibiotic exposure is a measure that reduces the number of false positives by excluding patients who receive a brief course; however, there is a trade-off between sensitivity and specificity. Constructing receiver-operating characteristic curves helps to identify the amount of treatment with the best combination of sensitivity and specificity. For example, acceptable identification of infections after cesarean section was achieved by requiring a criterion of at least 2 days of parenteral antibiotic administration (21). In that study, the sensitivity was 81% and the specificity was 95% compared with infections identified by NNIS surveillance. Quantitative inpatient antibiotic exposure is useful for identifying infections in coronary artery bypass surgery patients (22). Receiver-operating characteristic curves were used to demonstrate that patients with infections were best identified as those who received postoperative antibiotics for at least 9 days, excluding the first postoperative day. This criterion included both oral and parenteral antibiotics and ignored gaps in administration. This approach has two important implications for surveillance systems: It allows this mechanism to identify patients readmitted for treatment.
of infection within 30 days of surgery, and automated programs to identify patients who meet this threshold are substantially easier to implement. The 9-day exposure cutoff resulted in greater sensitivity (approximately 90%) for identifying surgical site infections than conventional prospective surveillance (approximately 60%) conducted in the same hospitals. A disadvantage of the antibiotic threshold criterion is that it identifies events that are not surgical site infections, including problematic wounds that do not meet the HICPAC criteria for infection, other types of hospital infections, and other long durations of antibiotic use.

Studies under way will determine the utility of this approach in a larger number of hospitals. Preliminary data from nine hospitals suggest that surveillance for antibiotic use provides useful information. For cesarean section, prospective comparison of a quantitative antibiotic exposure threshold to conventional prospective NNIS surveillance and International Classification of Diseases, 9th Revision (ICD-9), discharge diagnosis codes indicates that antibiotic surveillance has considerably better sensitivity (89%) than either NNIS surveillance (32%) or coded discharge diagnoses (47%). This difference was consistent across hospitals (23).

Quantitative thresholds for antibiotic exposure should be chosen individually for specific surgical procedures, since the value for cesarean section (2 days) differs from that for coronary artery bypass grafting (9 days) and there may be no useful threshold for some procedures. These values may also need to be reassessed as medical practice evolves. It will be important to understand the discrepancies between the results of formal NNIS surveillance and antibiotic surveillance. In some cases, patients who receive more than the threshold duration of antibiotic therapy appear to have clinically relevant infectious illness, such as fever and incisional cellulitis with no drainage.

**Postdischarge Surveillance for Surgical Site Infection**

Because most infections become manifest after discharge and many patients with infections never return to the hospital where the surgery was performed (5), traditional inpatient surveillance methods are not sufficient. In addition, conventional methods for postdischarge surveillance, including surgeon questionnaires, are highly inaccurate, with both low sensitivity and specificity.

Information about postdischarge care is available in office-based electronic medical records of coded diagnoses, procedures, tests, and treatments from the automated billing and pharmacy dispensing data maintained by most HMOs and many insurers. Pharmacy dispensing information is typically available for insured patients who have a pharmacy benefit. Together, these automated data elements identified >99% of postdischarge infections that occurred after a mixed group of nonobstetric surgical procedures (5). This high sensitivity came at the cost of low specificity (many false positives requiring manual review of medical records).

Recursive partitioning, logistic regression modeling, and bootstrap methods have made it possible to preserve good sensitivity while improving specificity by combining automated data from inpatient and ambulatory sources. The resulting algorithms use these automated data to assign to each patient an estimated probability for postoperative infection. These probabilities of infection, based on postoperative events that indicate infection has occurred, must be distinguished from predictions based on personal risk factors such as diabetes or obesity or on characteristics of the procedures such as the duration of surgery.

Choosing a lower probability threshold results in higher sensitivity and lower specificity, whereas a higher threshold improves specificity at the expense of sensitivity. For example, using automated data from both HMOs and ambulatory medical records permitted a sensitivity of 74% and a specificity of 98%, for a predictive value positive of 48%. A higher sensitivity, 92%, was achieved at the expense of lowering the specificity to 92%, for a predictive value positive of 21% (Figure) (24).

This work has been extended to surveillance for inpatient and postdischarge surgical site infections following coronary artery bypass surgery in five hospitals (25). That study found that HMO data alone identified 73% of 168 infections and hospital data alone identified 49% of the same infections. Separate algorithms have been developed to identify postpartum infections occurring after discharge (26).

The utility of automated data sources might be improved in several ways: 1) A procedure-specific algorithm will likely perform better than a general one. 2) Algorithms can be improved to further reduce the number of false positives (e.g., by excluding codes for infection that occur on the same day as a surgical procedure or for antibiotics dispensed before the second postoperative day). 3) These algorithms should be made robust enough for general use by including all ICD-9 and Current Procedural Terminology codes that might be used for surgical site infections.

**Figure.** Performance of various methods for detection of postdischarge surgical site infections for 4,086 nonobstetric surgical procedures with no inpatient infection. Lines represent fitted receiver operating characteristic (ROC) curves for three logistic regression models, which differ by data sources available for generating probabilities. Points represent performance of four different recursive partitioning models and data from patient and physician surveys. For analyses limited to hospital data and outpatient antibiotic (Abx) dispensing data, the logistic regression model had equivalent performance to classification trees at the points shown. The fitted ROC curve falls below this point because most procedures clustered around a few discrete probabilities and limited data points cause approximation of the ROC curve to be less accurate. The recursive partitioning high-cost model accepts 15 false-positives at the margin to capture one true infection; the low-cost model accepts 5 false-positives at the margin (24). (Figure originally published in Sands et al. Journal of Infectious Diseases 1999;179:434. Copyright 1999, University of Chicago Press. Reprinted with permission.)
Improved Case-Mix Adjustment Methods

As quality improvement and patient safety programs evolve, there are likely to be many more opportunities and incentives for comparing infection rates within and across institutions. However, such comparisons will require case-mix adjustment that accounts for coexisting illnesses, to avoid penalizing hospitals that care for patients at higher risk. As discussed, the NNIS risk index is based on the ASA score, which has several undesirable features. Although the ASA score has five possible values, the NNIS index collapses them into two levels so that all information about coexisting illness is summarized, in effect, as high or low. There is often little heterogeneity of ASA score in patients within a surgical procedure class, for instance, cesarean sections. In addition, the ASA score is subject to considerable interobserver variation, is not available for many ambulatory procedures, is usually not captured in automated form by hospital databases, and is not available in administrative or claims data systems.

As an alternative to the ASA score, the chronic disease score has been proposed to adjust data for coexisting illness in surgical patients. This score is based on the premise that dispensed drugs are markers for chronic coexisting illness; for example, dispensing of hypoglycemic agents strongly suggests the presence of diabetes. Approximately 24 conditions are represented in the chronic disease score, which is computed from ambulatory pharmacy dispensing information and can predict death and overall resource use (27-30). The chronic disease score has theoretical advantages over the ASA score: it can be computed automatically for the approximately 90% of the population that has prescription drug coverage, and it is completely objective. In its first application to a mixed group of surgical procedures, the chronic disease score performed at least as well as the ASA score (30). In addition, a modified chronic disease score, based on data for drugs dispensed on hospital admission, performed with substantially better sensitivity and specificity than the ASA score. The chronic disease score, based on admission medications, can also be computed by health-care facilities without the need for ambulatory drug-dispensing data.

The chronic disease score might be considered as a substitute when the ASA score is not available or as a supplement to the ASA score to provide better risk stratification. In addition, the chronic disease score might be modified to optimize its prediction of surgical site infections, rather than all causes of death and resource utilization. For example, data on psychotropic drugs, which are important contributors to the overall chronic disease score, might detract from the prediction of infection. Improved scoring systems will need to be developed through formal modeling programs applied to large, heterogeneous datasets.

Potential Uses of Electronic Data for Surgical Site Infection Surveillance

Electronic data have the potential to provide better information about infections while reducing the effort required to conduct surveillance. The outcome measures (e.g., quantitative antibiotic exposure or combinations of coded diagnoses) are meaningful, although they differ from the NNIS definition. The medical profession must decide whether a surveillance definition of surgical site infection might coexist with a clinical definition, with the understanding that the two serve related but different purposes (for example, the surveillance definition for influenza epidemics depends on hospitalizations with a coded diagnosis of pneumonia or influenza rather than virologically confirmed infections or specific clinical signs and symptoms).

Implementation of systems that use these data requires consensus on the part of the medical profession about outcome definitions, surveillance algorithms, and reporting standards. Even if consensus is reached, impediments will remain to the widespread adoption of electronic surveillance systems. The disparity in the electronic systems currently in use is one of these. While more sophisticated systems will permit better surveillance, most of the results described above depend on data elements such as drug dispensing information or financial claims data that are already available or are among the first to become automated. Thus, it will not be necessary to wait for fully automated medical records or more advanced hospital information systems. Although the costs of developing and validating systems based on electronic data are substantial, much of the development can be centralized, and validation need only be conducted in a few sites to establish generalizability. These reporting systems require a moderate investment by hospitals, HMOs, and insurers, most of which is the fixed cost for creating automated reporting functions. While some of this cost can be defrayed through the use of standard, shared computer code, this code usually must be customized to make it compatible with existing automated systems. Organizations that have electronic data typically create similar reports for other purposes and will not need new skills. In addition, the costs of maintaining and using the periodic reports that will constitute a new surveillance system are negligible.

Data sharing between hospitals, HMOs, and insurers is important, since very few single entities possess enough information to implement a self-sufficient surveillance system. Furthermore, in many locales, hospitals contract with several HMOs and insurers. In that case, HMOs and insurers must share information among themselves as well as with the hospitals, since no one hospital is likely to have enough patients to achieve the necessary precision. Data sharing will require development of systems that protect both patients’ confidentiality and the organizations’ proprietary interests.

If such surveillance becomes widely available, two types of uses might coexist. One would be to improve traditional prospective surveillance; for example, sensitivity of inpatient surveillance could be maintained with greatly reduced effort by restricting traditional (NNIS) review to the <10% of records that meet the quantitative screening criterion for antibiotic exposure. Similarly, for the postdischarge surveillance system, one could review as little as 2% of records (including ambulatory records in physicians’ offices) while greatly increasing the sensitivity of detection.

A second way to use these surveillance systems is to apply them to the entire surgical population, including patients or procedures that are not being evaluated because of resource constraints. Tracking the proportion of inpatients who exceed the antibiotic threshold or the number of patients who exceed a prespecified computed probability of surgical site infection after discharge might be sufficient, as long as that proportion is within agreed-upon limits. When the rates are below this limit, no further evaluation would be needed, since important problems in the delivery system are unlikely to have escaped detection. However, when the proportion or number exceeds...
the prespecified limit, more rigorous examination of the data would be triggered.

Electronically assisted surveillance for infections could be performed at modest expense by many organizations that have administrative claims and pharmacy data. These groups include the providers of care for most of the U.S. population, including essentially all HMO members, many of those with traditional indemnity insurance, Medicaid recipients, and most Medicare beneficiaries who have pharmacy benefits.

Supported in part by cooperative agreement UR8/CCU115079 from CDC.

Dr. Platt is professor of ambulatory care and prevention at Harvard Medical School, hospital epidemiologist at Brigham and Women’s Hospital, and director of research at Harvard Pilgrim Health Care, an HMO.

References


