Administrative Coding Data and Health Care–Associated Infections

Michael A. Jhung and Shailen N. Banerjee
Division of Healthcare Quality Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia

Surveillance for health care–associated infections (HAIs) using administrative data has received attention from health care epidemiologists searching for efficient means to track infections in their institutions. Several states are also considering electronic surveillance that incorporates administrative data as a means to satisfy an increasing demand for mandatory public reporting of HAIs. International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) discharge diagnosis codes have attributes that make them suitable for detecting HAIs; for example, they may facilitate automated surveillance, freeing up infection control personnel to perform other important tasks, such as staff education and outbreak investigation. However, controversy surrounds the appropriate use of ICD-9-CM data in detecting HAIs, and administrative coding data have been criticized for lacking elements necessary for surveillance. Administrative coding data are inappropriate as the sole means of HAI surveillance but may have value to the health care epidemiologist as a way to augment traditional methods.

Surveillance for health care–associated infections (HAIs), a fundamental responsibility of hospital epidemiology departments, is conducted largely by infection preventionists (IPs). Because it involves collection, analysis, interpretation, and dissemination of data, HAI surveillance can be extremely time, labor, and resource intensive. Administrative coding data have been proposed to support certain aspects of HAI surveillance to alleviate some of this resource burden. This review explores the potential roles for discharge diagnosis codes in HAI surveillance and identifies some of their principal limitations as surveillance instruments. It concludes with cautionary advice for applying these data to HAI detection efforts.

Administrative data are receiving consideration as a means to track HAIs in the United States, partly in response to recent state and federal policy initiatives [1–5]. Several states currently require hospitals to report HAIs; as of 2008, 31 mandate public reporting of HAI rates in some capacity [6]. This number has increased markedly since 2002, when only 4 states had legislation requiring health care facilities to disclose HAI rates [4]. Although facilities in 19 states currently mandate use of the National Healthcare Safety Network, which is a Centers for Disease Control and Prevention (CDC) system for active surveillance of HAIs, a number of states either do not specify a data source or explicitly identify administrative data as the source for HAI rates [1]. In addition, the Deficit Reduction Act of 2005 (DRA) has directed the Centers for Medicare and Medicaid Services to reduce payments to hospitals for conditions associated with complications of care that stem from certain hospital-acquired conditions. Beginning in October 2008, provisions of the DRA stipulate that hospitals no longer receive additional payment for 10 selected conditions, 3 of which are HAIs (table 1) [7]. Currently, cases are identified using administrative discharge data.

Nearly all hospitals already use International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) discharge diagnosis codes; many facilities also routinely generate other administrative data (eg, pharmacy dispensing records and claims data). The application of administrative data to surveillance purposes could therefore often be accomplished at relatively modest additional effort and expense.

Although primarily used for claims processing, ICD-9-CM...
codes have attributes that make their use appealing for tracking HAIs. First, hospital discharge data are available and readily accessible electronically; their incorporation into HAI surveillance efforts may thus free up IPs for other important tasks, such as staff education and outbreak investigation. Second, discharge databases use well-described terms that have been available since the 1950s and are adaptable to the changing needs of constituents via periodic updates. Finally, nearly all discharge databases follow Uniform Hospital Discharge Data Set standards and provide data using the Uniform Bill 92 format. The common format ensures that, once coding has occurred, data elements remain standard across facilities and regions and over time.

**LIMITATIONS OF ADMINISTRATIVE DATA**

Despite these attributes, there has been only modest use of ICD-9-CM codes in surveillance for HAIs, for which an IP-centered approach remains a widely accepted standard. The extent to which discharge data may enhance traditional HAI detection depends on many factors. Users should recognize that administrative data, particularly ICD-9-CM codes, were not designed for surveillance purposes. Instead, their primary role is in remuneration; this creates a major limitation of ICD-9-CM surveillance for HAIs.

**Diagnosis code lists can be artificially abbreviated.** After discharge, medical coders examine patient records for documented medical conditions and create a list of ICD-9-CM codes, which are then arranged into diagnosis related groups by proprietary software. Consistent and complete provider documentation is recommended and coders are instructed not to interpret abnormal findings (eg, laboratory, radiographic, pathological, or other diagnostic results) without provider documentation of their clinical significance [8].

A patient’s list of discharge diagnoses consists of a single primary diagnosis and zero to several secondary diagnoses. Although hospitals often retain a complete list of diagnoses, the maximum number reported to the Centers for Medicare and Medicaid Services is 9; the number collected by the National Center for Health Statistics currently is 7, and the number collected by the Agency for Healthcare Research and Quality is 15. Because hospitals use discharge diagnosis lists to petition for reimbursement of costs related to medical care, diagnoses can be listed in order of expected remuneration, rather than in order of clinical importance. Thus, ICD-9-CM codes that represent important clinical conditions but are associated with low reimbursement can be dropped from a list. Conditions identified by ICD-9-CM supplemental codes can be particularly vulnerable to such exclusion, because they are used only to modify other diagnoses and thus cannot be principal diagnoses. For example, the ICD-9-CM code V09.0 is defined as an infection with microorganisms resistant to penicillins (methicillin-resistant *Staphylococcus aureus* [MRSA] infection) and may only be listed as a secondary diagnosis. V09.0 was used by Elixhauser and Steiner [9] to identify 368,600 nationwide hospitalizations for MRSA infections in 2005. In separate analyses of hospital systems in Illinois and Utah, V09.0 was found among the first 9 listed diagnoses in only 52% of discharges with confirmed MRSA infection and was found beyond the 15th diagnosis in 15% of such discharges [10] (CDC, unpublished data). Extensive changes in reimbursement structure, such as those posed by the DRA, may also affect the ordering of discharge diagnoses and in an unpredictable manner.

**Diagnosis codes may not correspond directly to clinical syndromes.** In addition, clinically related entities are not linked on lists of ICD-9-CM codes; thus, syndromes that require >1 code to be completely described may be difficult to identify from a list of discharge diagnoses. A coder may describe an *S. aureus* surgical site infection (SSI), for example, by using the *S. aureus* organism code (041.11) and an SSI code (eg, 998.5 for postoperative infection). These 2 codes would appear as independent entries among a patient’s discharge diagnoses. If this patient also carried a diagnosis of *S. aureus* cellulitis, the patient’s list of ICD-9-CM codes could also include 681.00 (cellulitis and abscess of the finger, unspecified). From these 3 codes alone, it is not possible to discern whether the organism code for *S. aureus* (041.11) is used to describe the SSI, the cellulitis, or both. It is therefore difficult, using ICD-9-CM codes alone, to determine estimates of conditions that require multiple codes for their accurate description (eg, an infection at a specific site caused by a particular pathogen).

**Clinicians and coders do not always speak the same language.** Discharge diagnosis data have been criticized for having poor sensitivity, having poor positive predictive value (PPV), and lacking key elements necessary for surveillance of many important HAIs [5, 11–18]. For example, Sherman et al [11] assessed sensitivity and PPV for detection of central line–associated bloodstream infection, catheter-associated urinary tract infection, ventilator-associated pneumonia, and SSI using ICD-9-CM codes compared with targeted active surveillance by IPs. For ICD-9-CM codes, overall sensitivity was 61% and PPV was only 20%. Stevenson et al [12] reported similarly low PPV values for central line–associated bloodstream infection.

**Table 1. Health Care–Associated Infections Included in the List of 10 Hospital-Acquired Conditions for Potential Reduced Payment Effective 1 October 2008 under the Deficit Reduction Act of 2005**

<table>
<thead>
<tr>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catheter-associated urinary tract infection</td>
</tr>
<tr>
<td>Vascular catheter-associated bloodstream infection</td>
</tr>
<tr>
<td>Surgical site infection, including:</td>
</tr>
<tr>
<td>Mediastinitis after coronary artery bypass graft surgery</td>
</tr>
<tr>
<td>Orthopedic surgery of the shoulder and elbow and spinal fusion</td>
</tr>
<tr>
<td>Bariatric surgery for morbid obesity</td>
</tr>
</tbody>
</table>

950 • CID 2009:49 (15 September) • HEALTHCARE EPIDEMIOLOGY
they may be used to complement traditional surveillance provided by the available evidence. However, we propose several reliance on discharge data alone for HAI surveillance is not supported by the available evidence. Therefore, to address the limitations of ICD-9-CM codes, we propose several strategies to improve their accuracy and utility.

**ADDRESSING LIMITATIONS**

Several recent studies have estimated the burden of HAIs in US hospitals using discharge databases. These studies have used various methods and have derived different results [9, 13–18, 22–31]. Results of these studies may be used to propose more effective strategies for using administrative data in HAI surveillance (table 2).

**HAI NATIONAL BURDEN ESTIMATES**

The annual burden of HAIs in the United States is difficult to estimate, because no single source of representative data is currently available. Because ICD-9-CM discharge codes standardize clinical vocabulary across facilities, they may be well suited to estimate national or regional HAI burden. Furthermore, if used consistently over time, administrative data can be appropriate for estimating national trends, a task for which accuracy may not be as important as consistency in measurement. For surveillance purposes, a system with low sensitivity can be useful in monitoring trends if sensitivity remains constant over time [32].

**Use a database with as many diagnosis fields as possible.**

One way to improve sensitivity is to ensure that every ICD-9-CM code assigned to a patient at discharge is available for collection. This point is illustrated by comparing estimates of *S. aureus* and *Clostridium difficile* infections derived using 2 different databases. The Nationwide Inpatient Sample (NIS) is a discharge database maintained by the Agency for Healthcare Research and Quality as part of their Healthcare Cost and Utilization Project. The NIS is a nationally representative database of US hospital inpatient stays that collects data on up to 15 diagnoses for each patient sampled [33]. The National Hospital Discharge Survey (NHDS), conducted annually by the CDC, is also a nationally representative database of US hospital inpatient stays, but it currently only includes data for up to 7 diagnoses for each patient [34]. In separate studies, national estimates of *S. aureus*–related and *C. difficile*–related hospitalizations using the NIS were substantially higher than estimates obtained using the NHDS [24, 30, 35] (CDC, unpublished).
data); this may be due, in part, to the greater number of diagnosis fields available in the NIS (figure 1). The NHDS recognizes this limitation and intends to increase the number of diagnosis fields available for each hospitalization from 7 to 15 in 2010. Thus, although not feasible for all patients (eg, those with dozens of diagnoses), it may be prudent to use a database with as many diagnosis fields as possible for surveillance purposes.

**Use multiple codes and combinations of codes.** Another method that may improve sensitivity of HAI detection is using multiple ICD-9-CM codes, instead of a single code, to identify a clinical syndrome. For example, surveillance for *S. aureus* sepsis can be conducted using the single code for this specific syndrome (038.11), but because of variability in clinician documentation and coder interpretation, it is reasonable to suspect that additional true cases of *S. aureus* sepsis are represented by other codes in the discharge diagnosis list. Thus, including ICD-9-CM codes suggestive of *S. aureus* sepsis in addition to 038.11 may improve system sensitivity. In fact, Ollendorf et al [36] showed an increase of 12.3% in sensitivity when including combinations of codes that would likely indicate sepsis. Similarly, Moro and Marsillo [25] demonstrated an 11% increase in sensitivity of SSI detection when adding nonspecific codes suggesting postoperative complication to a list of SSI-specific codes, although this improvement only raised sensitivity to 21%.

**Identify a range of estimates.** Recognizing the difficulty with which diagnoses of *S. aureus* infections are captured by a single ICD-9-CM code, we developed a hierarchical algorithm to identify 3 estimates of *S. aureus*–related discharge rates—a conservative estimate, a moderate estimate, and a liberal estimate—using combinations of codes [35]. Conservative estimates were based on diagnosis codes for *S. aureus* sepsis (038.11) and *S. aureus* pneumonia (482.41), the *S. aureus* organism code (041.11), and multiple codes for skin conditions; moderate and liberal estimates included additional codes. Applying this algorithm to the NIS database, we found the moderate and liberal estimates of *S. aureus* sepsis to contain 21% and 54% more *S. aureus*–related discharges for sepsis, respectively, than the conservative estimate.

**Validate results against other estimates from nonadministrative sources.** To assess the relative accuracy of these 3 estimates for *S. aureus* sepsis, we compared results to an estimate of disease obtained from a nonadministrative surveillance system. Using an active, population-based surveillance system, Klevens et al [37] estimated the US burden of initial cases of invasive MRSA infections to be 94,360 in 2005 [37]. Klevens et al defined a case of invasive MRSA infection as isolation of MRSA from a normally sterile body site, including blood, cerebrospinal fluid, pleural fluid, pericardial fluid, peritoneal fluid, joint or synovial fluid, bone, internal body site, or other normally sterile site. Although the resulting estimate of invasive MRSA infection would therefore encompass more than sepsis, most invasive infections were identified from blood isolates. Our results (table 3) suggest that a liberal estimate incorporating multiple ICD-9-CM codes may yield the best approximation of the burden of *S. aureus* or MRSA sepsis.

**FACILITY-LEVEL HAI IDENTIFICATION**

Because of variation in coding practices and clinician documentation, use of ICD-9-CM codes may not be uniform across all facilities; therefore, case identification at the facility level should not be undertaken with a goal of interfacility comparison. Instead, when administrative data can be combined with other electronic information sources within a facility, discrete algorithms can be used, perhaps as an initial strategy, to uncover patients at higher risk for HAI, who could then receive a more thorough evaluation.

**Augment detection by using additional administrative or electronic data.** Combining coding data with other electronic sources (eg, pharmacy dispensing information and microbiology databases) can improve sensitivity and PPV for HAI detection. Much of the work performed in this area has focused on SSI surveillance, which seems to be well suited to a system combining discharge codes with pharmacy data, and central line–associated bloodstream infection surveillance, which is suited to detection via a combination of ICD-9-CM codes and electronic microbiology data.

In a series of ongoing studies, Sands et al [16, 17], Yokoe et al [18], and Platt et al [27] have assessed the use of discharge diagnoses and postoperative antimicrobial exposure to identify SSIs. When using procedure-specific algorithms, including both ICD-9-CM codes and postoperative antimicrobial exposure, they found substantial improvements in sensitivity and PPV. For example, in a multicenter study of 8790 coronary artery bypass graft procedures, Yokoe et al [18] found that sensitivity improved from 54% using diagnosis codes alone to 93% when adding antimicrobial exposure. Interestingly, sensitivity for antimicrobial exposure alone was 91%, supporting the authors’ claim that including diagnosis codes may be unnecessary to maintain acceptable sensitivity. Baker et al [38] found comparable results in a study of endometritis after cesarean delivery, concluding that a combination of coding and pharmacy data would become the case-finding method of choice at their institution.

Numerous studies demonstrate the usefulness of conducting surveillance for HAIs using electronic microbiology databases [39–43]. Several assessments have also been performed of laboratory-based systems that incorporate administrative data to
assist with detection of bloodstream infections [13, 29], SSIs [26, 28], and HAIs in general [44–47]. Many of these designs involve sophisticated computer algorithms that fully automate HAI identification. For example, Trick et al [29] developed an algorithm using microbiology, pharmacy, and administrative data that displayed 81% sensitivity and 81% PPV to detect central venous catheter–related bloodstream infections. Bellini et al [13] also devised an automated surveillance system, incorporating microbiological and administrative data that, under optimal conditions, was 78%–98% sensitive in detecting health care–associated bloodstream infections.

For some facilities, a combination of discharge codes and microbiology reports may be helpful for SSI identification, as well. Chalfine et al [26] developed a computer-assisted surveillance system for SSI detection using administrative and microbiology data that was 84% sensitive and nearly 100% specific [26]. Indeed, the accurate classification of patients without SSI might be the greatest benefit of adding microbiology data to surveillance using administrative data. Postoperative antimicrobial exposure may identify patients without true SSI who are receiving antimicrobials for other reasons. This liability is largely overcome by the addition of microbiology data, which accurately separates postoperative patients with SSI from postoperative patients without SSI. Spolaore et al [28] corroborate this assertion by showing that PPV improves from 70% to 97% when microbiology data are added to discharge codes in an SSI surveillance system [28]. Several studies have also described improvements in surveillance parameters for HAI detection in

For some facilities, a combination of discharge codes and microbiology reports may be helpful for SSI identification, as well. Chalfine et al [26] developed a computer-assisted surveillance system for SSI detection using administrative and microbiology data that was 84% sensitive and nearly 100% specific [26]. Indeed, the accurate classification of patients without SSI might be the greatest benefit of adding microbiology data to surveillance using administrative data. Postoperative antimicrobial exposure may identify patients without true SSI who are receiving antimicrobials for other reasons. This liability is largely overcome by the addition of microbiology data, which accurately separates postoperative patients with SSI from postoperative patients without SSI. Spolaore et al [28] corroborate this assertion by showing that PPV improves from 70% to 97% when microbiology data are added to discharge codes in an SSI surveillance system [28]. Several studies have also described improvements in surveillance parameters for HAI detection in
general when using multiple databases instead of a single source [44–46].

CONCLUSION

Analysis of currently available administrative data does not appear to be capable of functioning as the sole means of conducting surveillance for HAIs. The Healthcare Infection Control Practices Advisory Committee recognizes the value of discharge databases and encourages their use as a data source for case finding but not as the sole source for surveillance [4]. Although administrative data may augment traditional surveillance methods, any application of administrative data to HAI detection requires a thorough understanding of limitations and a willingness to interpret surveillance results with caution. One particular concern is that administrative data may be used for interfacility comparison of HAI rates, but differences in medical practice, medical record keeping, and coding make such comparisons unfounded.

If provider documentation, coding practice, and reimbursement incentives remain consistent over time, the most appropriate use of discharge codes may be as a means to track trends in national estimates of HAIs. However, because there are factors that could jeopardize the consistency of these measures (eg, the DRA and introduction of the Present on Admission indicator for diagnoses), it will be important to evaluate further the suitability of this application of discharge data over time. For case identification at the facility or local level, combining administrative data with other electronic data seems to be prudent, when possible, if used to contribute to a multipart surveillance system wherein electronic data are used to identify patients at increased risk for HAI who merit subsequent direct evaluation by IPs.

Acknowledgments

We thank Scott Fridkin, L. Clifford McDonald, and Daniel Pollock for their assistance with the preparation of the manuscript.

Potential conflicts of interest. All authors: no conflicts.

References