Automated Surveillance for Healthcare-Associated Infections: Opportunities for Improvement

Maaike S. M. van Mourik,¹ Annet Troelstra,¹ Wouter W. van Solinge,² Karel G. M. Moons,³ and Marc J. M. Bonten¹,³

¹Department of Medical Microbiology, ²Department of Clinical Chemistry and Hematology, and ³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, the Netherlands

Surveillance of healthcare-associated infections is a cornerstone of infection prevention programs, and reporting of infection rates is increasingly required. Traditionally, surveillance is based on manual medical records review; however, this is very labor intensive and vulnerable to misclassification. Existing electronic surveillance systems based on classification algorithms using microbiology results, antibiotic use data, and/or discharge codes have increased the efficiency and completeness of surveillance by preselecting high-risk patients for manual review. However, shifting to electronic surveillance using multivariable prediction models based on available clinical patient data will allow for even more efficient detection of infection. With ongoing developments in healthcare information technology, implementation of the latter surveillance systems will become increasingly feasible. As with current predominantly manual methods, several challenges remain, such as completeness of postdischarge surveillance and adequate adjustment for underlying patient characteristics, especially for comparison of healthcare-associated infection rates across institutions.

Keywords. healthcare-associated infection; surveillance; prediction; electronic; methodology.

Healthcare-associated infections (HAIs) pose a considerable burden on medical care worldwide. Approximately 1.7 million people develop an HAI in US hospitals each year, and 7% of patients admitted to European hospitals are affected by an HAI on any given day [1, 2]. Over the past decades, HAIs have more often been considered preventable complications of medical care, and reporting of infection rates has become increasingly important not only for healthcare providers, but also to payers and the public. The ongoing extension of pay-for-performance programs and (mandatory) public reporting has increased the potential impact of HAI rates on healthcare reimbursement and public opinion [3–5]. Given the anticipated consequences of reported HAI rates, the accuracy of surveillance results is more of a concern than ever before.

Surveillance and feedback of HAI rates has been a cornerstone of infection prevention programs since the 1980s, as participation in surveillance programs was associated with reduced HAI incidence [6–8]. Surveillance programs require reliable HAI rates, delivered in a timely manner and as efficiently as possible. Traditionally, infection control professionals manually review medical records of all patients at risk for the specified HAI or conduct prospective inpatient surveillance. However, this is highly resource intensive and can therefore only be applied to limited patient populations. Moreover, case definitions usually leave room for subjective interpretation and the completeness of case finding depends on the effort put in by surveyors [9, 10]. Finally, differences between hospitals in how surveillance is implemented may affect reported infection rates. For instance, case finding limited to a single source of information (microbiology results, antibiotic prescriptions)
may lead to underreporting [11, 12], as does more thorough application of case definitions [13].

These drawbacks of traditional HAI surveillance methods have led to initiatives to support or replace manual surveillance, either prospectively or retrospectively, by automated data collection from electronic medical records of microbiology results or antibiotic dispensing and/or administrative data [14–30]. Electronic surveillance methods aim to distinguish patients with high and low likelihoods of having developed an HAI to limit manual records review to high-risk patients; this improves efficiency while the number of missed HAI cases is maintained at an acceptable level. These electronic systems can be divided into 2 categories (Table 1): classification algorithms that select high-risk patients based on the presence of indicators of infection and multivariable regression models that combine indicators of infection in a weighted formula to identify high-risk patients.

In this review, we will discuss that existing electronic surveillance systems based on classification algorithms are an improvement over manual surveillance but that limitations remain with respect to the balance between completeness of case finding (sensitivity) and efficiency. More sophisticated multivariable prediction models, which to date have found only few applications in this context, address these limitations and may generate the high-quality infection rates that are needed for (public) reporting and infection control research.

**CLASSIFICATION ALGORITHMS**

Electronic surveillance systems based on classification algorithms are analogous to a classification tree that classifies patients as low or high risk of having had an infection based on several consecutive dichotomous (yes/no) decision steps. In first instance, case finding is based on flagging patients with a certain dichotomous indicator of HAI (eg, positive microbiology culture results from the targeted infection site). In each subsequent step, classification of patients is further refined by applying a new “HAI indicator” such as the 30-day postoperative interval, presence of a specified device, or concurrent antimicrobial dispensing. Patients classified as high risk after these consecutive steps are then selected to undergo medical records review (Figure 1A) [18, 21, 25–27]. This method has been applied most often to surgical site infections (SSIs) [17–22] and catheter-associated bloodstream infections [24–28], but also to ventilator-associated pneumonia, urinary tract infections, or postpartum infections [28–30]. Most classification algorithms use 1 or 2 dichotomous indicators of infection stored in a structured format in hospital databases to flag high-risk patients; for instance, microbiology culture results, (threshold) antimicrobial exposure, or administrative data (discharge and procedure codes) [16–22, 24, 28–31]. Classification algorithms have also been used by payers (eg, insurance companies) to directly estimate HAI rates from claims data without manual confirmation and perform ranking of healthcare providers [31, 32].

As these classification algorithms aim to select, or rather exclude, patients who certainly did not have an HAI and do not require manual records review, the most important performance characteristics are sensitivity (high probability of detecting the truly affected patients) and negative predictive value, the probability that absence of a flag by the algorithm truly excludes HAIs. The efficiency of the system can be assessed by the positive predictive value, the probability that the presence of a flag by the algorithm in fact represents HAI, or alternatively, by the number needed to screen or review, which is the number of records that must be manually assessed to identify 1 HAI [14]. In the setting of payer-based surveillance, important performance characteristics of classification algorithms are precision of the estimated total number of HAI at the group level and ranking accuracy [16, 31].

**Performance of Existing Classification Algorithms**

Compared to manual surveillance, electronic classification algorithms have a high efficiency, and some show improved intrarater reliability and sensitivity, possibly due to increased consistency of classification of microbiology culture results and introduction of more systematic case-finding strategies, respectively [18, 21, 24, 27]. Table 2 lists performance characteristics of classifications algorithms for surveillance of HAI from recent studies and shows the gain in efficiency achieved.

However, the studies listed in Table 2 also illustrate some caveats of these classification algorithms. The fine balance between sensitivity and positive predictive value (efficiency) is exemplified by the SSI algorithms. In the algorithms with acceptable levels of sensitivity (>80%), true infections were only confirmed in 1 of 3–8 medical records flagged for review
The consequences of this restrictive case finding will depend on the type of HAI surveyed (bloodstream infection being the main exception) and clinical practice. Also, the use of discharge or procedure codes as a means to replace or support manual surveillance has revealed suboptimal results [34]. Although this readily available source of data may be an attractive alternative to costly manual surveillance, coding relies on an intermediate step of data interpretation where other priorities than those necessary for HAI surveillance may prevail.

In short, electronic classification algorithms have important advantages over manual HAI surveillance both in terms of

---

Figure 1. Schematic overview of electronic surveillance system structures. In traditional surveillance, all patients will undergo manual records review. A, Classification algorithm where case finding is based on the (unweighted) presence of 1 or more indicators of infection followed by subsequent (dichotomous) refinement of flagged patients (based on administrative data, device presence, other indicators of infection). High-risk patients can then undergo records review. B, Regression model where all sources of data are included in the application of a weighted regression formula. All patients with a predicted probability of infection above the specified threshold will undergo manual records review to confirm the infection. Gray: patient developed HAI; black: patient did not develop HAI. Sources of information circled in gray are not often used in the specified system. Sources of information with dashed lines are currently only available through natural language processing. Abbreviations: ADT, admission, discharge, and transfer database; HAI, healthcare-associated infection.
Table 2. Characteristics and Performance of Selected Classification Algorithms for the Detection of Healthcare-Associated Infections

<table>
<thead>
<tr>
<th>Study</th>
<th>Targeted HAI</th>
<th>No. (% HAI)</th>
<th>Setting</th>
<th>Case Finding Based on</th>
<th>Refinement Steps</th>
<th>Sensitivity (%)</th>
<th>PPV (%)</th>
<th>NNS</th>
<th>Records to Review (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trick et al, 2004 CLABSI</td>
<td>135 (35.6)</td>
<td>Inpatient</td>
<td>Microbiology (blood culture)</td>
<td>Timing, other cultures, antibiotic use</td>
<td>81</td>
<td>62</td>
<td>1.3</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Woeltje et al, 2008 CLABSI</td>
<td>771 (8.7)</td>
<td>ICU</td>
<td>Microbiology (blood culture)</td>
<td>Timing, CVC presence, other culture, fever</td>
<td>94</td>
<td>20</td>
<td>5.0</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>Woeltje et al, 2011 CLABSI</td>
<td>391 (22)</td>
<td>Inpatient</td>
<td>Microbiology (blood culture)</td>
<td>Timing, CVC presence, other culture, fever</td>
<td>95</td>
<td>90</td>
<td>1.1</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Pokorny et al, 2006a BSI, VAP, CAUTI</td>
<td>194 (18)</td>
<td>ICU</td>
<td>2 or more of: microbiology, antibiotics or discharge code</td>
<td>Timing</td>
<td>94</td>
<td>56</td>
<td>1.8</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Stamm and Bettacchi, 2012</td>
<td>141 HAI</td>
<td>ICU</td>
<td>Microbiology (NIM)</td>
<td>Timing, ADT</td>
<td>67</td>
<td>39</td>
<td>2.6</td>
<td>Unclear</td>
<td></td>
</tr>
<tr>
<td>Leth et al, 2010b UTI</td>
<td>1513 (3.2)</td>
<td>Inpatient + postdischarge</td>
<td>Microbiology and/or antibiotic</td>
<td>Timing</td>
<td>77</td>
<td>93</td>
<td>1.1</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>SSI</td>
<td>1513 (7.1)</td>
<td>Inpatient + postdischarge</td>
<td>Microbiology and/or antibiotic and/or discharge code and/or reoperation</td>
<td>Timing</td>
<td>72</td>
<td>71</td>
<td>1.4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Bolon et al, 2009c SSI</td>
<td>6322 (1.7)</td>
<td>Inpatient</td>
<td>Antimicrobial exposure, diagnosis codes (index + readmission)</td>
<td>Timing</td>
<td>~90</td>
<td>25–40</td>
<td>2.5–4</td>
<td>4.0–6.2</td>
<td></td>
</tr>
<tr>
<td>Yokoe et al, 2012c SSI</td>
<td>11159 (5.8)</td>
<td>Inpatient</td>
<td>Antimicrobial exposure, diagnosis codes (index + readmission)</td>
<td>Timing</td>
<td>89</td>
<td>18</td>
<td>5.6</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Song et al, 2008 SSI</td>
<td>1226 (5.9)</td>
<td>Inpatient + outpatient</td>
<td>Antimicrobial exposure, discharge diagnosis (index, follow-up)</td>
<td>Timing</td>
<td>82</td>
<td>13</td>
<td>7.7</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>Hollenbeak et al, 2012d</td>
<td>1066 (8.8)</td>
<td>Inpatient</td>
<td>Microbiology (NIM)</td>
<td>Timing, ADT</td>
<td>20</td>
<td>68</td>
<td>1.5</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>Gerbier et al, 2012 SSI</td>
<td>446 (8.5)</td>
<td>Inpatient</td>
<td>Antimicrobial exposure</td>
<td>Timing</td>
<td>68</td>
<td>34</td>
<td>2.9</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: ADT, admission, discharge, and transfer database; BSI, bloodstream infection; CAUTI, catheter-associated urinary tract infection; CLABSI, central line–associated bloodstream infection; HAI, healthcare-associated infection; ICU, intensive care unit; NIM, nosocomial infection marker; NNS, number needed to screen; PPV, positive predictive value; SSI, surgical site infection; UTI, urinary tract infection; VAP, ventilator-associated pneumonia.

Sources: [18, 20–23, 25–28, 30]. Reference standard was manual medical records review in routine surveillance unless specified otherwise.

- Reference standard based on 194 patients.
- Case finding for the reference standard includes patient questionnaire.
- Data based on extrapolation of medical records review in a random sample (both studies use very similar methodology).
- Reference standard from surgical quality improvement program.
efficiency and systematic application of case definitions, but suffer from vulnerability to restrictive case finding and the payoff between sensitivity and efficiency.

**MULTIVARIABLE REGRESSION MODELS**

An alternative approach is the use of multivariable regression (prediction) models that combine indicators of infection (or predictors) simultaneously—rather than consecutively as in classification algorithms—to discriminate between high- and low-risk patients and select patients for manual medical records review (Figure 1B) [17, 32, 35]. Given the patient’s observed data, the weighted regression formula estimates the probability of having had an infection; this can then be used to stratify patients as having a high or low risk of HAI based on a probability threshold selected by the user [16, 36].

This approach has been used to a limited extent by combining diagnosis and procedure codes extracted from claims databases with antimicrobial dispensing records [19, 32]. As can be expected, models combining the largest number of predictors performed better than more parsimonious models [17]. Furthermore, the use of clinical patient data for the surveillance of a broad group of HAIs in a simple model with 5 variables (antibiotic days, urine cultures performed, length of stay, wound cultures taken, and age) achieved reasonable performance and could reduce the surveillance workload to 33% while maintaining sensitivity above 80% [35].

In Table 3, we have used empirical data to demonstrate the advantage of multivariable regression models over the use of classification algorithms for the surveillance of drain-related meningitis, a complication of external cerebrospinal fluid drainage in neurosurgical patients that is complex and cumbersome to diagnose manually (details in [36]). The case definition does not necessitate isolation of micro-organisms from cerebrospinal fluid, accounting for the limited sensitivity of case finding based only on microbiology results (Table 3, numbers 1 and 2); as the population at risk has complex underlying disease, selection based on treatment with antibiotics leads to many false-positive predictions (Table 3, number 3). Broadening the algorithm to combine both microbiology results and antimicrobial exposure clearly improved sensitivity, but at the cost of lower efficiency (Table 3, number 5). The multivariable regression model combining all available predictors of drain-related meningitis simultaneously (Table 3, number 6) achieved near-perfect sensitivity and had the highest efficiency, with high positive predictive value and low number needed to screen. Moreover, by changing probability thresholds based on the number of missed cases deemed acceptable, users can modify the balance between sensitivity and specificity to their wishes (Table 3, number 7).

Although the input variables used by classification algorithms and regression models are similar, statistically weighted multivariable regression models have several advantages over classification algorithms:

<table>
<thead>
<tr>
<th>Model Structure</th>
<th>No. of Data Sources</th>
<th>Sensitivity (%)</th>
<th>PPV (%)</th>
<th>NNS</th>
<th>Records to Review (%)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Microbiological culture</td>
<td>2 (Device use, microbiology)</td>
<td>80/104 (77)</td>
<td>80/164 (49)</td>
<td>2.0</td>
</tr>
<tr>
<td>2</td>
<td>Microbiological culture for contaminant</td>
<td>3 (Device use, microbiology, pharmacy)</td>
<td>78/104 (75)</td>
<td>75/133 (56)</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>Any antimicrobial exposure (&gt;4 days)</td>
<td>2 (Device use, pharmacy)</td>
<td>100/104 (96)</td>
<td>100/331 (30)</td>
<td>3.3</td>
</tr>
<tr>
<td>4</td>
<td>Antimicrobial exposure to standard empiric regimen</td>
<td>2 (Device use, pharmacy)</td>
<td>81/104 (78)</td>
<td>81/153 (53)</td>
<td>1.9</td>
</tr>
<tr>
<td>5</td>
<td>2 or 4</td>
<td>3 (Device use, microbiology, pharmacy)</td>
<td>102/104 (98)</td>
<td>102/235 (43)</td>
<td>2.3</td>
</tr>
<tr>
<td>6</td>
<td>Regression modelb</td>
<td>4 (Device use, microbiology, pharmacy, clinical chemistry)</td>
<td>102/104 (98)</td>
<td>102/170 (60)</td>
<td>1.7</td>
</tr>
<tr>
<td>7</td>
<td>Regression model (higher threshold)c</td>
<td>4 (Device use, microbiology, pharmacy, clinical chemistry)</td>
<td>86/104 (83)</td>
<td>86/118 (73)</td>
<td>1.4</td>
</tr>
</tbody>
</table>

Data taken from [36]. All models presented require manual assessment of drain usage. If data were missing (models 1–4), results were classified as negative. Abbreviations: NNS, number needed to screen; PPV, positive predictive value.

a Percentage of records to review is defined as the number of records requiring manual review to confirm infection as a fraction of the entire population at risk that would require manual review if using traditional methods.

b Model threshold = 0.15.

c Model threshold = 0.3.
• Multiple indicators are simultaneously taken into account, rather than consecutively, making case finding less restrictive and allowing for greater variation in clinical presentation of infection and thus higher detection rates (sensitivity).
• The weighted formula ensures higher positive predictive values than would be achieved if case finding were broadened in a nonweighted fashion (by combining indicators in an “and/or” manner in classification algorithms).
• Prediction models offer flexibility with respect to the desired balance between sensitivity and efficiency. Increasing the predicted probability threshold above which patients are selected for manual review will lower the sensitivity of the model but increase efficiency and vice versa [36].
• Compared to previously developed regression models based on discharge and procedure codes extracted from claims data, a system that directly uses routinely collected clinical care data does not depend on data interpretation, such as assignment of discharge codes.
• Applications of prediction models may extend beyond their use as a stratification system. In particular, the number of infections at the group level can be estimated directly by summing predicted probabilities, thereby allowing estimation of infection rates without medical records review [36].

Developing Multivariable Prediction Models
Increasing adoption of electronic medical records make multivariable prediction models feasible for the surveillance of HAI. Models can be developed using standard epidemiological methods, and the number of data elements required may only be slightly larger than alternative classification algorithms with comparable sensitivity (Table 3) [37, 38]. So-called clinical data warehouses are particularly suitable for this purpose; the drain-related meningitis model was developed on the Utrecht Patient Oriented Database, a clinical data warehouse that links administrative and clinical databases [39]. Similar efforts and their application for HAI surveillance have been described for bloodstream infections [40], and other data mining efforts are currently being put toward the closely related surveillance of other medical complications (eg, adverse drug events) [41, 42]. With improvements in natural language processing, databases may in the future be enriched with increasingly detailed clinical information [42].

As with any attempt to develop surveillance methods, multivariable (logistic) regression models require data of a (sufficiently large) population at risk for the targeted HAI in which the candidate predictors have been measured and to which manual surveillance has been systematically and thoroughly applied by well-trained assessors (as reference or gold standard). Definition and choice of candidate predictors should be driven by clinical knowledge and feasibility of data collection. Because the available data elements are often limited to those collected during routine clinical care, nonrandom missing data will be common, and selection of predictors which are only rarely determined may complicate implementation of the model. Because missing values may also occur for indicators of infection that are commonly determined, methods such as multiple imputation of missing data may be needed to ensure reliable results during model development [43, 44]. Performance of the resulting model can be assessed by the model’s discrimination between infected and noninfected patients (eg, using the area under the receiver operating characteristic curve or, after introducing a threshold, the sensitivity, specificity, and predictive values) and model calibration, which quantifies the concordance between predicted and observed probabilities of infection [37]. Applying so-called internal validation strategies will assist in generating realistic estimates of model performance when applied in future populations [38].

After model development, external validation in patient populations independent from the development sample in time and/or place is essential to assess the generalizability of the model and provide insight in how model performance is affected by different patient characteristics, diagnostic practices, and prescription patterns. Finally, ongoing monitoring of model quality and model updating is needed as changes in patient characteristics and clinical practice may deteriorate model accuracy over time [38].

REMAINING CHALLENGES
Even if adoption of automated HAI surveillance based on regression models will be attainable in the (near) future, several issues need to be addressed to achieve the high-quality HAI rates needed by clinicians, policy makers, and researchers; some are specific to electronic surveillance, based both on classification algorithms and regression model methods, but many apply to all methods of surveillance. All methods of electronic surveillance require high-quality data as well as more advanced programming skills, statistical methods, and ongoing monitoring of data quality. A further challenge for all forms of electronic surveillance is assessment of device presence, such as Foley catheters and central lines, as this determines the population at risk for device-related HAIs. Depending on the functionality of the electronic medical record, the presence of devices is not always systematically recorded in a format that can easily be queried, necessitating manual assessment or proxy measures based on administrative data or, more promising, multivariable models [45–47]. Taking into account such functionality requirements during development and implementation of healthcare information systems would prevent this problem altogether.

Although methods for handling missing data in regression model development have been described extensively [43, 48], only limited work has been done how to apply developed
for the detection of ventilator-associated conditions; this has more objective and quantification in case ascertainment is, therefore, interpreted with care [50]. Moreover, all methods of surveillance described here rely on the quality and capacity of surveillance.

Because of the limitations of manual surveillance, both in terms of efficiency and reliability, research comparing alternative methods of surveillance is complicated by difficulties in obtaining a valid reference standard. Furthermore, partial and differential verification of the true infection status may bias results; therefore, reported performance characteristics must be interpreted with care [50]. Moreover, all methods of surveillance described here rely—at least partly—on manual surveillance to confirm the presence of an infection. Minimizing subjective interpretation in case ascertainment is, therefore, needed to further improve HAI surveillance. For instance, more objective and quantifiable parameters have been defined for the detection of ventilator-associated conditions; this has been proposed as an alternative for surveillance of ventilator-associated pneumonia, an infection that is notoriously difficult to diagnose [51]. With ongoing improvements in electronic algorithms and a shift toward measuring more objective outcomes, electronic surveillance may become the mainstay for some HAIs at least for the purpose of quality assessment.

Finally, if the HAI rates are to be used for benchmarking or comparisons across multiple institutions, accurate adjustment for underlying risk (case mix) is of vital importance. Development of more refined methods of case-mix adjustment is ongoing, and these efforts may benefit from the same improvements in healthcare information technology as the surveillance methods themselves [52].

CONCLUSIONS

The use of electronic classification algorithms for surveillance of healthcare-associated infections has increased reliability compared to the traditional surveillance through manual medical records review. However, the use of more sophisticated multivariable prediction or regression models is likely to further improve in sensitivity and efficiency of surveillance programs. Some important challenges such as postdischarge surveillance, quantification of device utilization, and case-mix adjustment are partly common to all methods of surveillance and need to be addressed in the future. With ongoing improvements in healthcare information technology, implementation of regression models will become feasible for widespread use and improve the quality and capacity of surveillance.

Notes

Acknowledgments. We thank the department of hospital hygiene and infection control for providing data on drain-related meningitis resulting from their ongoing surveillance efforts.

Financial support. This work was supported by the University Medical Center Utrecht (an Alexandre Suermann stipendium to M. S. M. vM.) and the Netherlands Organization for Scientific Research (NWO VICI grant 918.76.611 to M. J. M. B. and NWO VICI and TOP grants 9120.8004 and 918.10.615 to K. G. M. M.).

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

4. Association for Professionals in Infection Control and Epidemiology. HAI reporting laws and regulations. APIC 2011. Available at:
20. Song X, Cosgrove SE, Pass MA, Perl TM. Using hospital claim data to
10. Mayer J, Greene T, Howell J, et al. Agreement in classifying blood-
92
16. Platt R, Yokoe DS, Sands KE. Automated methods for surveillance of
2009
41. Bates DW, Evans RS, Murff H, Stetson PD, Pizziferri L, Hripcsak G.
42. Murff HJ, FitzHenry F, Matheny ME, et al. Automated identification of
36. van Mourik MS, Moons KG, van Solinge WW, et al. Automated detec-
9. Klompas M. Interobserver variability in ventilator-associated pneumo-
10. Mayer J, Greene T, Howell J, et al. Agreement in classifying blood-
stream infections among multiple reviewers conducting surveillance.


