Decision Making During Healthcare-Associated Infection Surveillance: A Rationale for Automation

William E. Trick

1Cook County Health and Hospitals System; and 2Rush University Medical Center, Chicago, Illinois

Attention to healthcare-associated infections has increased, in part due to legislative mandates for monitoring infections and federal payment policies. Current systems, which rely on considerable human involvement in finding and interpreting whether clinical events represent infection, can lead to biased institutional rankings. Relying on individuals employed by reporting institutions in an environment in which reporting healthcare-associated infections can be associated with punitive consequences is suboptimal. Cognitive psychology literature leads us to expect underreporting, economic theory suggests that underreporting will increase over time, and statistical theory indicates that there is a ceiling on reliability. With current systems, infection rates are likely to decline without meaningful improvement in practices. Fortunately, development of reliable and objective definitions and automated processes for infection determination has accelerated. Transition to such systems will be challenging; however, the result will be more valid interfacility comparisons.

Keywords: surveillance; algorithms; decision making; cognition; infection.

Increasingly, healthcare-associated infection (HAI) surveillance is mandated and rates are made publicly available, which drives scrutiny of the reliability and accuracy of infection determinations. In concert with mandatory reporting requirements, to include the Centers for Medicare Services' partnership with the Centers for Disease Control and Prevention's (CDC) National Healthcare Safety Network (NHSN) for submitting HAI data, the visibility of infection prevention programs has increased dramatically. Such visibility and fiscal consequences lead to increased awareness of infection rates by personnel at local healthcare facilities (eg, facility administrators). In the United States, the most prominent system for HAI reporting is NHSN, which serves as the HAI reporting system for >5000 US healthcare facilities [1], and an increasing number of rehabilitation centers and long-term acute care facilities. NHSN was the successor to the National Nosocomial Infection Surveillance (NNIS) system, and differs in that participation in the NNIS system was entirely voluntary and the system protected the confidentiality of participating facilities, whereas for many healthcare facilities, participation in NHSN is in response to legislative mandates [2, 3].

Consistent with the principles of establishing cut-points for diagnostic testing, there is an inverse association between sensitivity and specificity. These cut-points are determined inclusive of factors other than a predetermined mathematical relationship. For example, highly sensitive cut-points are employed to avoid missing a potentially catastrophic clinical event, such as pulmonary embolus or myocardial infarction. When reporting to NHSN, infection preventionists undoubtedly feel pressure to make highly specific determinations; the consequences of missing a device-associated infection impacts the infection preventionist and their institution less than overcalling an infectious event. Also, it is unlikely that clinical events not determined to be an
HAI will be scrutinized by individuals at the local facilities, whereas events classified as infection may undergo confirmatory review. Indirect evidence that infections are underreported during routine surveillance comes from the literature reporting infection rates from randomized and before–after clinical trials, in which rates often are much higher than the pooled mean infection rates reported to NHSN. In randomized trials, due to sample size considerations, there may be an interest in using sensitive rather than specific interpretations.

Even in circumstances in which infections are not being reported to derive institution-specific infection rates, the reliability of infection determinations has not been good [4–7]. To address concerns about the accuracy of reporting, some states audit infection determinations [8, 9]; however, the prospect of systematically auditing infection determinations is daunting. More sustainable improvements in reliability will be realized through NHSN’s commitment to develop infection definitions that increase objectivity, or through transitions to definitions more readily automated.

Prior papers have discussed the advantages of transitioning to more automated and perhaps algorithmic methods of infection detection [10, 11]. In this paper, I discuss the theory and evidence underpinning suboptimal reliability of reporting of HAIs and the reasons why underreporting is likely to increase over time, particularly when used for performance measurement tied to reimbursement, rather than internal quality improvement initiatives. Most of the theories and evidence are borrowed from other disciplines, but I believe these theories can be generalized to human behavior around reporting infection rates. I present options for modified surveillance methods that are being considered or implemented to improve the validity of interfacility comparisons.

**HUMAN FRAILTY**

The context and affective inputs of situations can have a profound impact on decision making, particularly in situations of uncertainty. Because of the uncertainty in classifying clinical events [12–14], attributing a patient’s condition to HAI can be ambiguous. Despite the best intentions of infection preventionists, poor reliability and decreasing infection rates should be expected based on prior experimental evidence reported by cognitive psychologists and economists.

**Cognitive Psychology**

Infection preventionists often are directly involved in implementing the interventions they believe to be effective. Because they also perform outcome assessments, we should expect, based on theory and evidence from research settings, that even sham interventions would artifically reduce infection rates. Because of the well-known biases inherent in human decision making, a fundamental principle underlying clinical trials is that participants, caregivers, outcomes assessors, and treatment allocation be blinded to participants’ status as either exposed to the intervention or usual care [15]. Without such safeguards, the efficacy of interventions can be overstated [15]. Although infection surveillance is an operational rather than a research endeavor, there still is the need for decision making in the context of considerable uncertainty.

Affective inputs (eg, anxiety, complaisance, and fear) influence information processing and decision making, particularly during reasoning in situations of uncertainty [16, 17]. Such inputs, although useful for rapid decision making, can result in illusory interpretations of information in positive or self-protective directions [18]. Affective inputs that arise from the context in which the decision is being made have been termed the “framing effect” [19]. Currently, infection preventionists make determinations in a climate in which they are beset with messages that HAIs are preventable, and they may be concerned about confrontation from clinical directors charged with lowering publicly reported complication rates [20, 21]. Such messages are being conveyed in parallel with implementation of interventions popularly considered to be more effective than empirically shown in rigorous trials [22]. Because beliefs about the effectiveness of such interventions likely are exaggerated, infection preventionists review potential events with the preconception that since they promoted effective processes, the likelihood of HAI is low. In statistical terms, infection preventionists likely review potential infectious events with a pretest probability distorted to unrealistically low levels. The lower the baseline probability of infection, the less likely it is that additional clinical information will shift the probability beyond the infection preventionist’s dichotomous yes/no threshold for HAI determination.

Not only is processing existing information biased by affective inputs, but acquisition and retention of new information or information incongruent with one’s preformed opinion can be biased, for example, through search or attribution bias [23]. Reviewers who at baseline believe that a clinical event does not represent an HAI will be less likely to search for controvertible evidence, and more likely to explain away evidence of infection as inconsequential. This might be manifest in searching through the clinician’s narrative for an opinion that counters the presence of infection. One criterion for a laboratory-confirmed bloodstream infection due to a common skin commensal is the presence of chills. In the absence of a documented temperature elevation, infection preventionists may not be motivated to search the clinical narrative for documentation of chills.

**Iterated Public Goods State**

It is unpleasant to recognize that there is the possibility of calculated underreporting of infections in the context of performance...
measurement. Examples of such behavior outside of HAI surveillance are documented in experiments categorized as public goods games. Public goods games can be designed to simulate situations in which individuals decide whether to behave altruistically for the benefit of a group or make selfish decisions that provide immediate personal gains but harm the group [24–27]. Such games are designed to identify patterns in human behavior during situations in which individuals derive the greatest benefit if they behave selfishly. If every participant behaves altruistically, the reward for all individuals exceeds their contribution. An iterative game is one in which the simulation is repeated and behavior observed over time and rewards and/or punishments can be incorporated. Infection surveillance is analogous to an iterated public goods game in that infection preventionists are expected to behave altruistically for the benefit of a larger group (their future patient population and patients at other facilities by establishing accurate benchmarks), and there is longitudinal reporting of infection rates. Infection preventionists who underreport infections degrade the value of surveillance for the entire community; however, as individuals, they benefit from their lack of cooperation—low infection rates are celebrated.

Unfortunately, simulated public goods games have consistently found individuals who protect their self-interest, and in some simulations, only a minority of individuals behave altruistically [26]. In iterated public goods games, altruistic behavior diminishes over time as participants realize that others exploit the opportunity to behave selfishly. Although to date, there is no evidence that gaming the system is systematic [28], it is possible that over time, calculated underreporting may occur.

The conditions of public goods games can be manipulated to sustain altruistic behavior either through voluntary participation [24, 25, 27] (such as the CDC’s prior NNIS system) or structuring the system to reward altruistic behavior or punish selfish behavior [26]. Paradoxically, as currently structured, altruistic behavior is punished when a facility reports a high HAI rate.

STRUCTURAL LIMITS ON RELIABILITY

In addition to cognitive biases and potential gaming of the system, there are structural factors built into the current NHSN surveillance system. First, the infection preventionist is required to make a dichotomous (ie, yes or no) interpretation of a complex clinical scenario. Second, with subjective criteria included in HAI definitions, reliable case finding will always be a challenge.

Inherent Uncertainty

For both clinicians and infection preventionists, there is uncertainty in determining whether a clinical event represents an HAI [4]. Because there is uncertainty between the boundaries of “infection = definite yes” or “infection = definite no,” a more appropriate scale would provide latitude to express uncertainty, such as an ordinal scale (eg, strongly disagree; disagree; neutral; agree; strongly agree). When the range of responses for an item that conceptually is not dichotomously scaled are collapsed to fewer options, reliability can be compromised [29, 30]. Two theories for degradation in reliability when response options are limited are as follows: individuals have different thresholds for categorizing events, and small differences in estimated probabilities can be amplified (Figure 1). Infection preventionists have different thresholds for infection determinations, with some requiring a high degree of certainty before categorizing events as HAIs (ie, low-sensitivity, high-specificity raters) [4].

Over time, as infection preventionists encounter resistance at categorizing infections, thresholds for infection determinations may increase, resulting in enhanced specificity at the cost of reduced sensitivity.

Despite the theoretical advantage of expanding the scale for infection determinations, it is impractical to ask infection preventionists to estimate the likelihood of infection using either an interval or ordinal scale. Also, dichotomous determinations are valuable for purposes of clear communication and evaluation of events; however, we should recognize that the cost of such a scale is that there is a ceiling on reliability. A real-world example of a common dichotomous categorization is expressing approval or disapproval through a thumbs-up or thumbs-down gesture. As raters, we want to expand our options, which manifests as 2 thumbs up for very good or a noncommittal rotation of the hand to express mediocrity.

Subjectivity in Definitions

A more commonly recognized threat to reliability is the difficulty in consistently interpreting clinical evidence regarding infection determinations. In the NHSN system, there are examples of definitions that have substantial subjectivity or variability in documentation, such as the presence of suprapubic tenderness as a criterion for urinary tract infection, or whether a putative central line–associated bloodstream infection (CLABSI) was “not related to an infection at another site.” Also, NHSN allows for the physician or surgeon diagnosis of infection to be an acceptable criterion for an HAI. Although the wording is structured only as a criterion for the presence of an HAI, it is possible that infection preventionists use physician interpretation as evidence against infection.

Even seemingly objective definitions, such as whether the event occurred within a specified period of time (eg, present or incubating on admission), can be interpreted different ways in separate institutions [31]. In regard to the relatively simple task of assigning infections to the unit of origin, questions arise, such as when does the clock start for an admission (eg,
emergency room door, time the emergency room requests an inpatient bed, patient arrival on the inpatient care unit, or time of admission orders? To its credit, NHSN recognized these challenges and has begun to implement changes in its operations and definitions [32].

STRATEGIES TO IMPROVE RELIABILITY

Current Manual Methods

There are options short of full automation to improve the reliability of current methods for categorizing clinical events as HAIs (Table 1). Although some strategies may be prohibitively costly or impractical, others already have been implemented or are under development, such as external audits [8, 9], modification of definitions to use more objective criteria [33], and improving training methods and materials. Partnerships with external organizations are being formed to improve the accuracy of reporting, and such collaborations should be beneficial. State health departments are one such potential partner and are increasingly involved in validation efforts. In 2011, 14 states validated CLABSI data in 2011, 5 states validated surgical site infection reporting, and 4 states validated catheter-associated urinary tract infection reporting [1]. The Centers for Medicare and Medicaid Services also is beginning a validation review during fiscal year 2013 [34]. Although such efforts should improve reliability, the costs likely are substantial and such programs may be difficult to sustain without ongoing federal support, and such audits will be insufficient to address all concerns. Recent increases in the number of NHSN-participating facilities undoubtedly presents challenges for training and providing personalized assistance for difficult determinations.

An extreme option to minimize bias would be to share clinical records across facilities or to a centralized location for interpretation of events deemed to be at high probability for an HAI; such determinations could be made independent from local influences. Unfortunately, such a system would require substantial investments in electronic medical record capacity to allow for sharing clinical information to remote auditors, a scenario that is likely many years from being realized.

More imminent is that NHSN is working to improve the reliability of their system through changes in definitions or operations designed to maximize the reliability of assimilating complex clinical data. These changes focus on more objective categorization of HAIs, such as a 2-day calendar rule for determining community vs healthcare acquisition, which replaced the “present or incubating at the time of admission” rule, and eliminating the sometimes ambiguous definition of implants for surgical site infection [32]. NHSN continues to systematically train infection preventionists and guide external audits of HAI determinations. Most importantly, they are actively progressing with objective infection determinations that allow for automation, such as the substitution of ventilator-associated event detection to replace ventilator-associated pneumonia [33, 35].

---

Figure 1.  Graphic representation of the 2 reasons reliability can be compromised when a measure that is conceptually an interval scale is collapsed to 2 (dichotomous) categories. A, Both infection preventionists (IPs) independently determine that a clinical event had a 75% likelihood of being a healthcare-associated infection (HAI). Despite having an equivalent interpretation of the likelihood of infection, IP A and IP B have different thresholds for HAI determination. IP A has a lower sensitivity threshold and will categorize the episode as not being an HAI, whereas IP B would record the event as an HAI. B, Three IPs have minor differences in their interpretation of the likelihood that clinical events represent an HAI; probability estimates are clustered within a few percentage points. Although they have the same cut-point for an HAI, the small differences in their probability determinations are amplified by requiring a yes (event = 1.0) or no (event = 0) determination. Abbreviations: HAI, healthcare-associated infection; IP, infection preventionist.
Automated Methods

Fully automating infection detection through use of discrete data elements captured from the electronic medical record would obviate many of the reliability problems that plague systems reliant on human interpretation. Automated systems would not be affected by cognitive biases, intentionally manipulating infection rates would be more difficult and likely transparent, and modeling strategies could be used to estimate probabilities of events rather than yes/no determinations. Estimation of probabilities through modeling of clinical parameters should avoid the dramatically increased infection rates likely to be observed with algorithmic infection detection [5]; rather than binary algorithmic categorization, overall rates would be generated by summing estimated probabilities [36, 37].

For some HAIs, notably ventilator-associated complications, urinary tract infection, or CLABSIs, fully automated infection detection is likely to be realized by some institutions [38, 39], and recognized by agencies responsible for surveillance systems. Although there has been substantial progress in using algorithms for surgical site infection case finding [40], there has been less progress toward full automation of infection detection [37].

Methods developed for automated detection of infection should focus on inclusion of data elements that represent a direct result of infection or place a patient at high risk of HAI, that are computable, and that are consistently available during clinical care (eg, clinical culture results, length of stay, temperature, oxygenation requirements for ventilated patients, device use). Less ideal is incorporation of data elements subject to variability in clinical behaviors (eg, antimicrobial use) or those not routinely reported (eg, device cultures or differential time to positivity for blood cultures). Use of administrative (International Classification of Diseases, Ninth Revision) codes has not performed well as a method to generate accurate infection rates [41].

The challenges of automating infection determinations are numerous and in some instances complicated, but not insoluble. Foremost of these challenges is to build processes to record data in computable formats, and to build informatics capacity to aggregate clinical data and run either algorithms for yes/no determinations or models to estimate and sum probabilities. Also, there needs to be acceptance among the infection prevention community and clinicians that fully automated systems

---

**Table 1. Summary of the Hazards to Producing Reliable Estimates of Healthcare-Associated Infections, and Possible Solutions to Improve the Current System**

<table>
<thead>
<tr>
<th>Hazard</th>
<th>Example</th>
<th>Possible Solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive bias</td>
<td>An IP implements a campaign to reduce central line–associated bloodstream infections. The belief that infections will be prevented biases against identifying infections.</td>
<td>Audit determinations discordant with automated algorithmic results. Impartiality may only be possible through outside expert review. Minimize consequences of high infection rates for surveillance personnel. Large infection control departments could separate surveillance and prevention roles. Set realistic goals for HAI prevention. Expecting zero events likely influences determinations. Rather than rank-order facilities, use stringent methods for identifying outliers.</td>
</tr>
<tr>
<td>Intentional underreporting</td>
<td>An IP perceives job insecurity after reporting high infection rates. Subsequently, they intentionally report lower rates.</td>
<td>Routinely audit a sample of determinations, especially in facilities in which discrepancies with automated methods are common.</td>
</tr>
<tr>
<td>Subjectivity in definitions</td>
<td>For the same patient, IPs disagree as to whether a patient has urinary symptoms.</td>
<td>Migrate to more objective definitions. NHSN recently has modified HAI definitions to improve reliability.</td>
</tr>
<tr>
<td>Dichotomous determinations</td>
<td>For the same patient, IPs have different thresholds for what constitutes an infection.</td>
<td>Record interpretations on an ordinal or interval scale, such as a scale that allows a probability of 0.5 for infection. Not practical for manual methods, but relevant to automated methods in which probability estimates can be derived from models.</td>
</tr>
<tr>
<td>Attentivity</td>
<td>Attention to a hospital outbreak of tuberculosis leads to cursory surveillance, artificially lowering infection rates.</td>
<td>Staff infection control programs to account for outbreaks, vacations, and illnesses. Automated systems less susceptible to variations in staffing or workload.</td>
</tr>
</tbody>
</table>

For all hazards, automating infection determination should improve reliability.

Abbreviations: HAI, healthcare-associated infection; IP, infection preventionist; NHSN, National Healthcare Safety Network.
provide value by generating reliable comparisons between institutions. Because objective automated systems might miss the nuanced interpretation of a skilled clinician, there will be episodes that contradict the judgment of clinicians, which will lead to frustration. Substituting probability estimates rather than dichotomous determinations may mitigate such frustration; however, relying on estimated probabilities for internal quality improvement initiatives may be awkward and require manual confirmation of episodes that have a relatively high probability of infection.

As definitions migrate to more objective criteria and electronic medical records evolve to more completely capture clinical events in discrete computable formats, moving to automated systems to track infections will provide a better measure of the effectiveness of our infection prevention interventions and guide us to those most likely to protect our patients from HAIs.

Notes

Acknowledgments. My thanks go to Michael Lin, MD, MPH, for his thoughtful review and edits.

Potential conflicts of interest. Author certifies no potential conflicts of interest.
The author has 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


